



FACULTAD DE ESTUDIOS ESTADÍSTICOS

**MÁSTER EN MINERÍA DE DATOS E
INTELIGENCIA DE NEGOCIOS**

Curso 2019/2020

Trabajo de Fin de Máster

***TÍTULO:* Análisis y predicción de los estudiantes
extranjeros en universidades españolas.**

***Alumno:* Jaime Nicolás Ordóñez Torres**

***Tutor:* Juana María Alonso Revenga**

Julio de 2020



UNIVERSIDAD COMPLUTENSE
MADRID

Índice General

1. Introducción	1
1.1. Justificación del proyecto	1
1.2. Contexto	2
2. Fuentes de Datos	3
2.1. Origen de los datos	3
2.2. Naturaleza de los datos	5
3. Objetivos y Metodología	8
3.1. Objetivos del Proyecto	8
3.2. Metodología	9
4. Métodos estadísticos aplicados	11
4.1. Modelos Descriptivos	11
4.1.1 Análisis de Correspondencias Simple	11
4.2. Modelos de series Univariantes	12
4.2.1 Conocimientos previos	12
4.2.2 Modelos basados en métodos de suavizado	14
4.2.3 Modelos estacionarios	15
4.2.4 Modelos no estacionarios	16
4.3. Metodología Box-Jenkins	17
4.4. Modelos de Regresión lineal	20
4.4.1 Regresión LASSO	20
4.4.2 Regresión PLS	21
5. Desarrollo del Análisis	22
5.1. Análisis y descripción de las Variables	23
5.1.1 Sexo	24
5.1.2 Edad	25
5.1.3 Región de procedencia	25
5.1.4 País de procedencia	27

5.1.5 Comunidad Autónoma de destino	28
5.1.6 Nivel de estudios	30
5.1.7 Tipo de institución Universitaria.....	33
5.1.8 Precios	36
5.8.1.1 Precios de estudios universitarios	36
5.8.1.2 Precios de alquiler	39
5.2. Análisis de correspondencias	41
5.2.1 Relación entre Edad y Región de procedencia.....	41
5.2.2 Relación entre Región de procedencia y CCAA de destino	45
5.3. Modelos de Regresión.....	47
5.3.1 Regresión LASSO.....	47
5.3.2 Regresión PLS.....	50
5.4. Modelos de Predicción.....	54
5.4.1 Análisis de la Serie temporal.....	54
5.4.2 Aplicación de Modelos de Predicción.....	57
5.4.3 Comparación y elección del mejor modelo.....	63
5.4.4 Predicción de estudiantes extranjeros en 2020 – ARIMA ajustado	65
5.4.5 Predicción de estudiantes extranjeros en 2020 – Holt-Winters	66
5.4.6 Combinación de modelos ARIMA ajustado y Holt Winters	67
6. Conclusiones	69
6.1. Trabajo Futuro.....	70
Bibliografía	71
Anexos	73
A. Código SAS.....	73
A.1. Análisis de Correspondencias Simple	73
A.1.1. Análisis entre Edad y Región de Procedencia.....	73
A.1.2. Análisis entre Región de procedencia y CCAA de destino.....	75
A.2. Regresión LASSO	77
A.3. Regresión PLS.....	78
A.4. Análisis y predicción de Serie temporal.....	79
B. Tablas de datos	85

Índice de Figuras

Figura 1. Esquema asociativo de fuentes de datos	4
Figura 2. Proceso de adecuación de modelos ARIMA	19
Figura 3. Universo y desagregación de datos utilizados	22
Figura 4. Evolución de estudiantes extranjeros desde 1995.....	23
Figura 5. Evolución y distribución de estudiantes extranjeros por Sexo	24
Figura 6. Distribución de estudiantes extranjeros en España por Sexo	24
Figura 7. Distribución de estudiantes extranjeros por Edad.....	25
Figura 8. Estudiantes extranjeros en España por Región de procedencia	26
Figura 9. Evolución de estudiantes por Región de procedencia desde 2010	26
Figura 10. País de procedencia de la mayor cantidad de estudiantes extranjeros.....	27
Figura 11. Distribución de estudiantes extranjeros por CCAA en 2019.....	28
Figura 12. Ciudades con más del 2% del total de estudiantes extranjeros.....	29
Figura 13. Evolución de extranjeros con estancia por estudios por ciudad de destino	29
Figura 14. Distribución de estudiantes extranjeros por Nivel de estudios	30
Figura 15. Estudiantes extranjeros por Nivel de estudios y por sexo	31
Figura 16. Estudiantes extranjeros por Nivel de estudios y CCAA de destino.....	32
Figura 17. Distribución de estudiantes extranjeros por tipo de Universidad (2018).....	33
Figura 18. Cantidad de Universidades por Comunidad Autónoma	36
Figura 19. Precios públicos para españoles y comunitarios por CCAA y nivel	37
Figura 20. Precios públicos para estudiantes extranjeros por CCAA y nivel	38
Figura 21. Comparativa de precios para estudiantes comunitarios vs extranjeros.....	39
Figura 22. Precio de alquiler mensual de habitación en piso compartido por CCAA	40
Figura 23. Precio de alquiler mensual de habitación en piso compartido por ciudad.....	40
Figura 24. Descomposición de inercia y chi-cuadrado - Edad vs Región.....	41
Figura 25. Mapa de calor de variables Edad y Región de procedencia	42

Figura 26. Plano factorial de variables Edad y Región de procedencia.....	43
Figura 27. Representación y contribuciones del perfil columna.....	44
Figura 28. Representación y contribuciones del perfil fila	44
Figura 29. Descomposición de inercia y chi-cuadrado Región vs CCAA.....	45
Figura 30. Representación del mapa de calor CCAA vs Región	46
Figura 31. Plano factorial de variables CCAA y Región de procedencia	47
Figura 32. Correlación lineal de las variables	48
Figura 33. Modelo seleccionado aplicando Regresión LASSO	49
Figura 34. Número de factores extraídos aplicando PLS.....	51
Figura 35. Estimadores de parámetros del modelo de Regresión PLS	52
Figura 36. Representación de la serie temporal entre 1993 y 2019	54
Figura 37. Comportamiento trimestral de la serie entre 2013 y 2016.....	55
Figura 38. Representación de la descomposición estacional de la serie	56
Figura 39. Serie original suavizada y predicciones de Holt-Winters	58
Figura 40. Representación del la ACF y PACF de la serie	59
Figura 41. Resultados de la serie diferenciada estacionalmente	60
Figura 42. Representación de resultados del modelo ARIMA ajustado	61
Figura 43. Resultados del Modelo ARIMA (2,0,0)(0,1,1) ₄	61
Figura 44. Factores autorregresivos y de media móvil del modelo	62
Figura 45. Representación de predicción del modelo ARIMA ajustado.....	63
Figura 46. Representación de predicciones versus datos reales.....	64
Figura 47. Representación de predicción 2020 con Arima ajustado	65
Figura 48. Representación de predicción 2020 con Holt-Winters	66
Figura 49. Representación de predicción de modelos y combinación de modelos.....	67
Figura 50. Contexto general de estudiantes extranjeros en España en 2020.....	68

Índice de Tablas

Tabla 1. Niveles de la clasificación CINE incluidos en la educación terciaria.....	6
Tabla 2. Países incluidos en cada región de procedencia.....	6
Tabla 3. Ranking de países con más estudiantes extranjeros en España.....	27
Tabla 4. Extranjeros matriculados por tipo de centro y nivel de estudios	34
Tabla 5. Estudiantes extranjeros de Máster en principales Universidades de España	35
Tabla 6. Tabla de contingencia entre variables edad y región de procedencia	41
Tabla 7. Contribuciones al estadístico Chi-cuadrado total de Edad vs Región.....	42
Tabla 8. Tabla de contingencia de Región de procedencia y CCAA de destino.....	45
Tabla 9. Variables tomadas en cuenta para la regresión LASSO.....	48
Tabla 10. Resultados del modelo seleccionado por regresión LASSO.....	50
Tabla 11. Variables tomadas en cuenta para regresión PLS	51
Tabla 12. Variación explicada por los factores - PLS.....	52
Tabla 13. Estimadores de parámetros de regresión PLS	53
Tabla 14. Principales estadísticos estacionales de la serie	55
Tabla 15. Suavizado de Holt Winters, método multiplicativo	57
Tabla 16. Predicciones de estudiantes extranjeros en España, método Holt-Winters.....	57
Tabla 17. Estadísticos de ajuste del suavizado Holt-Winters.....	58
Tabla 18. Residuales y comprobación de autocorrelación de la serie.....	59
Tabla 19. Predicciones de modelo ARIMA ajustado.....	62
Tabla 20. Predicciones para el 2019 del modelo ARIMA ajustado	63
Tabla 21. Comparación del error y medidas de bondad de los modelos.....	64
Tabla 22. Predicción trimestral 2020 aplicando Arima ajustado	65
Tabla 23. Predicción trimestral 2020 aplicando Holst-Winters	66
Tabla 24. Predicciones para 2020 y combinación de modelos	67

Capítulo 1

1. Introducción

1.1. Justificación

En un mundo cada vez más globalizado, las personas continuamente optan por estudiar en el extranjero buscando nuevas oportunidades más allá de sus fronteras para desarrollar plenamente sus talentos. La movilización de las personas engloba importantes beneficios, pero así también costos y riesgos tanto para los individuos, las instituciones y los países¹.

Por este motivo y en base al continuo aumento de la movilidad internacional con fines educativos y siendo España uno de los principales destinos en todo el mundo para estudiantes principalmente de educación terciaria, este TFM pretende realizar un análisis del contexto actual, observando la influencia y relación entre distintas variables como el sexo, edad, tipo de estudios, procedencia, etc. en la elección de España como destino de estudios, para lo cual se aplicarán técnicas de análisis de correspondencias y modelos de regresión lineal. También, a partir del análisis de una serie temporal, se ajustarán distintos modelos para predecir el número total de estudiantes que requieren autorización de estancia por estudios que llegarán al país a realizar sus estudios en el año 2020.

Este trabajo de investigación pretende ser un aporte para confirmar que efectivamente existe un continuo aumento de estudiantes extranjeros en España cada año y resultará de gran interés tomando en cuenta que la movilización de estudiantes extranjeros influye en temas importantes como la migración, la influencia en la sociedad española, su mercado laboral, transporte, precios de alquiler, etc. y además será de gran utilidad para el ámbito educativo, principalmente el Universitario, teniendo así información para la toma de decisiones sobre su oferta educativa y los beneficios o riesgos que implica la llegada de estudiantes desde el exterior.

¹ UNESCO. (2019). Informe de Seguimiento de la Educación en el Mundo. Recuperado de <https://en.unesco.org/>

1.2. Contexto

La proyección internacional de las universidades es un tema de interés estratégico tanto por el aumento en su prestigio, como por la recaudación extra de ingresos que aportan los estudiantes extranjeros. Por ejemplo, en 2016, se estima que los estudiantes internacionales y sus familias aportaron cerca de 39 400 millones de dólares a la economía de los Estados Unidos de América. En varios países asiáticos como Japón, la educación superior se abre a estudiantes extranjeros para que las instituciones permanezcan abiertas. Por todo esto, los países procuran armonizar cada vez más las normas y los mecanismos de garantía de calidad a nivel bilateral, regional o mundial².

Según la UNESCO y su informe sobre la ciencia (2030), entre 2005 y 2013 el número de estudiantes internacionales pasó de 2,8 millones a 4,1 millones en todo el mundo. Por lo que cada día, son más los estudiantes que eligen estudiar en el extranjero. Entre los principales factores que toman en cuenta los estudiantes para decidir dónde estudiar, es importante mencionar que hay tres criterios claves que son: el costo de vida, el estilo de vida y el ocio. Tres criterios con los que España ha logrado tener una ventaja competitiva que atrae anualmente a muchos estudiantes de distintas partes del mundo.

Por tanto, no es casualidad que España sea el país de la Unión Europea que más estudiantes de Erasmus recibe desde el 2015. Según la Organización para la Cooperación y el Desarrollo Económicos (OCDE), las universidades españolas en el último año acogían a más de 75.000 estudiantes extranjeros que realizan programas de grado y de postgrado. El último informe International Migration Outlook de la OCDE publicado en 2018, refleja que el número de universitarios internacionales que llegaron a España aumentó un 56%, una cifra muy alta si la comparamos con la media europea que fue del 8%, siendo así el país que ha registrado un mayor aumento de estudiantes internacionales.

Parte de este gran aumento se debe a la eliminación de la burocracia para cursar una carrera de grado completa en el país, esta decisión, tomada en 2014 de quitar la Selectividad para alumnos internacionales ha podido provocar una especie de efecto llamada. En cualquier caso, la internacionalización de la universidad española es aún muy mejorable. El número de estudiantes extranjeros que estudian en a nivel de grados y postgrados en España representan solo el 3% del total, cuando en Reino Unido son el 18% y en Francia el 10%.

² UNESCO. (2019). Informe de Seguimiento de la Educación en el Mundo. Recuperado de <https://en.unesco.org/>

Capítulo 2

2. Fuentes de Datos

2.1. Origen de los datos

Para este estudio se han recogido datos procedentes de distintas fuentes gubernamentales, organizaciones internacionales, de asociaciones universitarias, empresas privadas, etc. Lo que ha significado un trabajo de exploración y explotación de la información muy importante para obtener los datos más acertados sobre el tema.

En primer lugar, se han obtenido estadísticas de algunas fuentes del Gobierno español, a través del portal de inmigración del Ministerio de inclusión, seguridad social y migraciones, se han descargado ficheros de los extranjeros con autorización de estancia por estudios en vigor. A partir de estos datos, se ha hecho una recopilación y limpieza de ficheros, mediante la unión de varios archivos se ha generado una serie temporal trimestral que será fundamental para nuestro proyecto. De igual manera, usando este mismo portal, se han obtenido datos de las principales variables que se incluyen en tablas de contingencias para distintos modelos estadísticos.

Dentro del mismo ámbito gubernamental, se ha trabajado con varios ficheros de datos procedentes de las Estadísticas e Indicadores Universitarios que pone a disposición en su web el Ministerio de Ciencia e Innovación. También, a través del Ministerio de Educación y formación profesional, se ha utilizado el Sistema Integrado de Información Universitaria (SIIU), siendo una plataforma de recogida, procesamiento, análisis y difusión de datos del Sistema Universitario Español.

Finalmente, se han empleado otras fuentes que han sido de gran utilidad principalmente para el análisis del contexto actual de los estudiantes extranjeros en España, fuentes de datos fiables de organizaciones internacionales como la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y la UNESCO, Organización de las Naciones Unidas para la

Educación, Ciencia y Cultura. Además de otras fuentes de distinto carácter como los informes universitarios de la CRUE Universidades Españolas (asociación sin ánimo de lucro formada por 76 universidades españolas públicas y privadas) y demás portales de información estadística que han aportado datos relevantes (Idealista, Fotocasa, etc.).

La recopilación total de datos de distintas fuentes mencionadas se visualiza a continuación en la siguiente figura asociativa.

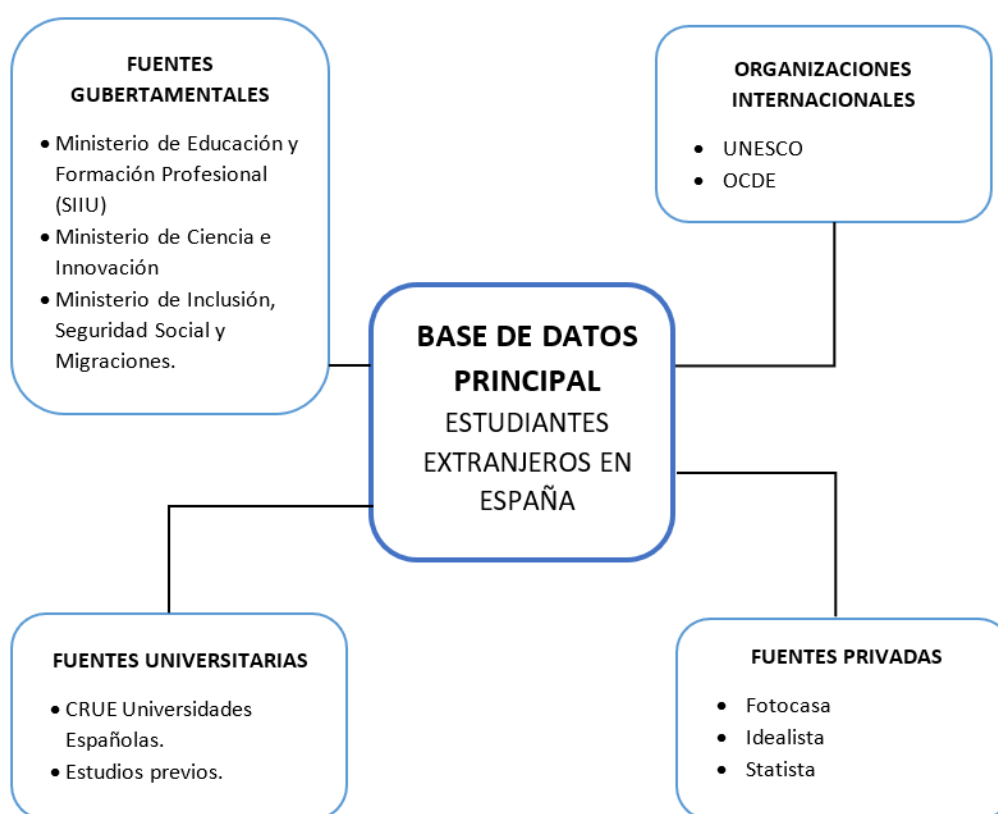


Figura 1: Esquema asociativo de fuentes de datos

2.2. Naturaleza de los datos

Al existir distintas modalidades de estudios principalmente en el ámbito de educación superior, tomando en cuenta programas de intercambio, de formación de corta o larga duración, etc. Habrá distintos perfiles de estudiantes extranjeros. Esto hace que aparezcan algunas limitaciones para nuestros datos y nuestro trabajo en general. Por lo que, en este análisis, principalmente se trabajará con el universo de estudiantes que solicitan una autorización de estancia por estudios.

La justificación para esto es simple, pues al ser la solicitud de estancia por estudios un trámite administrativo de relevancia, la información será más fiable y complementaria de la que ofrecen las estadísticas sobre enseñanza. Estos datos también permiten obtener una larga serie histórica que será muy útil para la predicción futura de estudiantes extranjeros. Es importante mencionar también dos posibles fuentes de inexactitud de los datos, el contabilizar más de una vez a una persona³ y la variación de países de cuyos ciudadanos precisan visados de estudios, eliminándose esa condición por ejemplo en el caso de países que en un año determinado hayan pasado a formar parte de la Unión Europea.

Para una mayor claridad en el concepto de estudiantes extranjeros, en este trabajo los definiremos como “todas aquellas personas que, residiendo en otro país y sin tener: condición de residencia española, residencia comunitaria o nacionalidad española, se desplazan al país para seguir un programa de estudios”. El ser estudiante extranjero, además, no implica siempre la obtención de un título universitario, como es el caso de los estudiantes de intercambio, sin embargo, habrá también otros estudios como posgrados o doctorados donde el objetivo principal es la obtención de una titulación.

Es importante también establecer que en este trabajo nos centraremos principalmente en los extranjeros que realizan estudios terciarios en España. Estos estudios se corresponden con los niveles de la clasificación CINE (Clasificación Internacional Normalizada de la Educación) de la UNESCO⁴, que incluyen los estudios que se detallan en la siguiente tabla.

³ Hasta el año 2008 no se tenía en cuenta el hecho de que una persona podía solicitar más de un visado de estudios el mismo año.

⁴ Clasificación Internacional Normalizada de la Educación CINE. (2011). Instituto de Estadística de la UNESCO.

Nivel CINE	Estudios que incluye
Nivel 5. Educación terciaria de ciclo corto	Educación de educación terciaria no suficiente para alcanzar un título de graduado licenciado.
Nivel 6. Grado en educación terciaria o nivel equivalente	Primer título, de 3 a 4 años, programas largos de primer título, de más de 4 años, y programa de segundo o siguiente título equivalente.
Nivel 7. Nivel de maestría, especialización o equivalente.	Programas largos de primer título, de al menos 5 años; programas de segundo o siguiente título equivalentes y programas de maestría, especialización o equivalente.
Nivel 8. Nivel de doctorado o equivalente	Programas que conducen directamente a un grado/título de doctorado.

Tabla 1: Niveles de la clasificación CINE incluidos en la educación terciaria

Dentro del universo de datos que hemos recopilado a través de las distintas fuentes mencionadas, usaremos las variables más relevantes que finalmente serán empleadas en el análisis y que se detallan brevemente a continuación:

- ❖ **Sexo:** Se diferenciará entre hombres y mujeres.
- ❖ **Edad:** Se usarán distintos rangos de edad de los estudiantes extranjeros; de 18 a 24 años; de 25 a 29 años; de 30 a 34 años; de 35 a 39 años; de 40 años o más.
- ❖ **Región de procedencia:** Se determinan seis regiones principales de procedencia, Resto de Europa, África, América del Norte, América Central y del Sur, Asia y Oceanía. A continuación, se ven los países que integran cada una:

Resto de Europa	Croacia, Albania, Andorra, Bielorrusia, Rusia, Serbia, Turquía, Ucrania, Otros Resto de Europa.
África	Angola, Argelia, Cabo Verde, Camerún, Costa de Marfil, Egipto, Etiopía, Ghana, Guinea, Guinea Ecuatorial, Libia, Marruecos, Mauritania, Nigeria, Rep. Dem. del Congo, Senegal, Túnez, Otros África
América del Norte	Canadá, Estados Unidos, México
América Central y del Sur	Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, Haití, Honduras, Nicaragua, Panamá, Paraguay, Perú, República Dominicana, Uruguay, Venezuela, Otros América, Central y del Sur
Asia	Arabia Saudí, Armenia, Azerbaiyán, Bangladés, China, Corea del Sur, Emiratos Árabes Unidos, Filipinas, Georgia, India, Indonesia, Irak, Irán, Israel, Japón, Jordania, Kazajistán, Kirguistán, Líbano, Malasia, Pakistán, Palestina, Siria, Tailandia, Taiwán, Uzbekistán, Vietnam, Otros Asia.
Oceanía	Australia, Nueva Zelanda, Otros Oceanía

Tabla 2: Países incluidos en cada región de procedencia.

- ❖ **País de procedencia:** Se identificará la nacionalidad de los estudiantes extranjeros según su país de procedencia.
- ❖ **Comunidad Autónoma de destino:** Comunidad Autónoma de destino de estudios de las personas extranjeras.
- ❖ **Nivel de Estudios:** Tipo de estudio para los que se solicita la autorización de estancia por estudios, diferenciando entre Grado, Posgrado y Doctorado.
- ❖ **Tipo de Institución Universitaria:** Se distinguirá principalmente entre Universidades Públicas y Privadas en donde realizan sus estudios las personas extranjeras en España, se tomará en cuenta también datos de centros universitarios privados (Escuelas de negocio, etc.).
- ❖ **Precios:** Se toma en cuenta el precio por crédito de los estudios universitarios según el nivel y según la institución universitaria. Se hace también un análisis del precio promedio de alquiler pagado por un estudiante extranjero, asumiendo que comparte piso.

Capítulo 3

3. Objetivos y Metodología

3.1. Objetivos del Proyecto

El objetivo principal de este proyecto es predecir a partir de una serie temporal, el número de estudiantes extranjeros que estudiarán en España en el año 2020 y tener un concepto claro del escenario y contexto de los estudiantes extranjeros en este país, analizando factores y variables que influyen en la elección de su destino y a la par observando la influencia que genera en España, su movilización.

Dentro de los objetivos específicos de este proyecto se han establecido los siguientes:

- ❖ Explotar la información disponible a través de gráficos que nos permitan tener una mejor comprensión del contexto actual de los estudiantes extranjeros.
- ❖ Realizar distintos análisis de correspondencias que nos permitan determinar las relaciones existentes entre las variables más importantes, aquellas que distinguen y caracterizan a los estudiantes extranjeros que realizan sus estudios en España.
- ❖ Aplicar modelos de regresión que sean una solución frente a la multicolinealidad de nuestros datos y nos permitan ver la correlación e influencia de las variables predictoras sobre el número de estudiantes extranjeros en España.
- ❖ Distinguir patrones y características de la evolución en el tiempo de los estudiantes extranjeros, a través del análisis de la serie temporal principal.
- ❖ Elección del mejor modelo de predicción de datos a partir de series temporales, a través de la comparación de los resultados con datos reales.

3.2. Metodología

La estructura y metodología que se aplicará en este trabajo tendrá dos partes bien identificadas. La primera se centrará en el análisis del contexto actual, mientras que la segunda en la predicción del número de estudiantes, para lo cual se emplearán distintas técnicas y modelos estadísticos utilizando el software SAS BASE.

En primer lugar y como parte del análisis del contexto actual de los estudiantes extranjeros en España, aplicaremos análisis factoriales de correspondencias simples. A través de esta primera metodología, será posible detectar relaciones existentes entre distintas variables que nos ayudará a tener una mejor perspectiva de las características y el perfil de los estudiantes extranjeros y conocer así su situación general en el país.

Para aplicar los análisis de correspondencias simples, crearemos varias tablas de contingencia, las mismas que se obtienen al cruzar dos variables nominales y repartir una población (o muestra) según el número de individuos que presentan una categoría de cada una de las variables. Como resultado, este método nos ayudará a representar la relación de las variables mediante un número reducido de dimensiones, buscando la menor pérdida de información posible.

Se buscará si las siguientes variables categóricas de los estudiantes extranjeros tienen relación o no, a partir de tablas de contingencias que cruzarán las siguientes variables:

- ❖ Región de procedencia vs Edad.
- ❖ CCAA de destino vs Región de procedencia.

Para robustecer el análisis de las relaciones entre variables, aplicaremos modelos de regresión que permitan ser una solución al inconveniente de la multicolinealidad de nuestros datos y mostrando la verdadera influencia de las distintas variables estudiadas en el número total de estudiantes extranjeros.

En segundo lugar, se logrará predecir el número de estudiantes extranjeros que llegarán a España a través del uso de series temporales, que no son más que la sucesión de valores observados en intervalos regulares en el tiempo. A partir del análisis de estas series, se aplicarán

técnicas o modelos de predicción para poder estimar el número de estudiantes que llegarán a futuro al país.

Se realizará un análisis descriptivo inicial de la serie, su descomposición estacional y se determinarán las principales características como la estacionalidad, tendencia, frecuencia, etc., para proceder a ajustar distintos modelos de series temporales a nuestros datos. En este caso se aplicará el modelo de suavizado Holt-Winters y luego un modelo ARIMA ajustado.

A partir de las predicciones obtenidas mediante estos dos métodos, podremos establecer qué modelo funciona mejor en nuestro caso, comparando los resultados de ambos versus los datos reales de los estudiantes extranjeros en España y observando qué modelo presenta menor error. Finalmente, a partir del mejor modelo elegido, se realizará una predicción del número estimado de estudiantes extranjeros que llegarán en 2020 a España, además, es importante destacar que el modelo puede ser actualizado con los datos recogidos cada año y utilizarse para predicciones en el futuro.

Capítulo 4

4. Métodos estadísticos aplicados

4.1 Modelos Descriptivos

4.1.1 Análisis de Correspondencias Simple

El **Análisis de Correspondencias Simple (ACS)** tiene como objetivo determinar la relación entre dos variables cualitativas a través de tablas de contingencias, representando los datos a partir de diagramas de dispersión de forma que sea posible detectar las relaciones existentes entre las variables.

Una **tabla de contingencia** se obtiene al cruzar dos variables nominales y repartir la muestra según el número de individuos que incluye cada categoría de las variables estudiadas. Las columnas de la tabla representan las modalidades o categorías de una variable, y el conjunto de filas representa los niveles de la otra variable a comparar. Las filas y columnas juegan papeles simétricos y se emplean de forma análoga (Alonso, 2019).

Para el estudio de la relación entre las variables que se cruzan en la Tabla de contingencia, se obtienen los denominados perfiles (fila y columna respectivamente), que representan las distribuciones condicionadas por cada una de las modalidades de la otra variable. Matricialmente se pueden calcular los perfiles fila y columna de la siguiente forma:

$$X_F = D_F^{-1}F \quad X_C = FD_C^{-1}$$

Donde $X_F (r \times c)$ y $X_C (r \times c)$ son los perfiles fila y columna respectivamente, $F = \{f_{ij}\} (r \times c)$ es la matriz de frecuencias relativas y $D_F = \text{diag}(f_i) (r \times r)$ y $D_C = \text{diag}(f_j) (c \times c)$ son las matrices diagonales que contienen las marginales de las filas y las columnas.

Para confirmar la relación o no de las variables, es importante hablar de la **hipótesis de independencia**. En este caso, consideramos que dos variables son independientes cuando el

valor que toma una variable no influye en la distribución de la otra. Aunque las gráficas brindan una idea general sobre la independencia, se debe recurrir a test estadísticos para sostener o rechazar la hipótesis de independencia.

Las **aportaciones al estadístico X^2** de los cruces de las categorías de las variables es una de las formas de obtener información relevante de relaciones, al indicarnos cuánto se alejan las frecuencias observadas de las esperadas. Si la aportación es alta, indica que las frecuencias observadas y esperadas difieren mucho, es decir, existe relación entre las categorías sea inversa o directa.

La Inercia es otro de los estadísticos que permiten evaluar la hipótesis de independencia. Se puede definir como una medida de la dispersión de la nube de puntos equiparable a la varianza de datos numéricos, pues representa la distancia X^2 de los perfiles al perfil medio ponderados por la masa de los perfiles (Alonso, 2019).

La representación en los planos factoriales es el último apartado importante para explicar el ACS, ya que permite representar los datos en espacios más reducidos y que sean representables en gráficos clásicos de dispersión, reduciendo así la pérdida de información.

Los métodos que se utilizarán para determinar cuántas dimensiones representar perdiendo la mínima información posible serán los siguientes:

- ❖ Tomar los dos o tres primeros ejes significativos, siempre que estos expliquen una variabilidad aceptable, mayor al 70%.
- ❖ Elegir las dimensiones cuya inercia sea superior a la media. Esto también se conoce como el *average rule*.

4.2. Modelos de series Univariantes

4.2.1 Conocimientos previos

En este trabajo, las predicciones obtenidas se harán en base al análisis de una serie temporal, por lo que resulta importante tener claro su definición y características. Una **Serie temporal** es el resultado de observar los valores de una variable a lo largo del tiempo en intervalos regulares. La teoría clásica establece que una serie de tiempo tiene cuatro componentes principales:

- ❖ **Tendencia (T_t):** el comportamiento general a largo plazo de la serie.
- ❖ **Componente estacional (S_t):** oscilaciones regulares que se producen de manera reconocible en el mismo periodo de tiempo y que se producen con un periodo igual o inferior a un año.
- ❖ **Componente cíclica (C_t):** refleja comportamientos recurrentes, se deben principalmente a ciclos en los que se repite el comportamiento, son difíciles de reconocer pues el periodo es difícil de identificar y en muchos casos variable.
- ❖ **Componente irregular (Z_t):** también conocido como aleatoriedad o ruido, son comportamientos irregulares por fluctuaciones causadas por sucesos impredecibles.

Estos cuatro componentes son los que permiten realizar el análisis de la serie, en general decimos que una serie puede ser expresada como suma de estos, de esta forma:

$$X_t = T_t + S_t + C_t + Z_t$$

O como un modelo multiplicativo, si la serie además del comportamiento estacional presenta tendencia o la componente estacional aumenta con el tiempo.

$$X_t = T_t \cdot S_t \cdot C_t \cdot Z_t$$

Siendo los procesos estacionarios los más utilizados al estudiar series temporales, es importante hablar de la herramienta básica con la que se cuenta para identificar un proceso estacionario. Este es el **coeficiente de autocorrelación**, definido por:

$$p_k = \frac{\text{cov}(x_t, x_{t-k})}{\sqrt{\text{var}(x_t) \text{var}(x_{t-k})}} = \frac{\gamma_k}{\gamma_0}, \quad k \geq 0$$

Donde $p_k = p_{-k}$ y $\gamma_k = \gamma_{-k}$ al tratarse de un proceso estacionario donde la correlación y covarianza de dos variables dependen solo del retardo.

De esta manera, se denomina **función de autocorrelación simple o ACF** a la representación de los coeficientes de autocorrelación en función al retardo, proporcionando la estructura de dependencia lineal de la serie. Sin embargo, genera influencias entre las observaciones que no aclara como influyen individualmente, para lo que se vuelve necesario aplicar la **función de autocorrelación parcial o PACF**, que para cada instante t y cada retardo k , toma un valor igual a la correlación entre z_t y z_{t+k} ajustada por el efecto de los retardos intermedios.

4.2.2 Modelos basados en métodos de alisado o suavizado

Estos modelos utilizan parámetros variables que son estimados, dando más importancia a los datos más recientes que a los antiguos. Así, se logra que los pesos que se les da a los valores de la serie decrezcan de forma exponencial y por lo tanto disminuyan la influencia de un valor según vamos hacia atrás en el tiempo. Su objetivo principal es hacer predicciones utilizando los datos de la serie eliminando las fluctuaciones aleatorias y tomando en cuenta solo la componente tendencia y estacionalidad (si la tiene). Se clasifican de la siguiente manera:

❖ Alisado Simple

Se utiliza cuando la serie no presenta tendencia creciente o decreciente. Se puede modelizar como: $X_t = L_t + Z_t$. Donde L es la tendencia nivelada y Z vendría ser el ruido o componente irregular de la serie. El cálculo de L_t se realiza en función del parámetro α , cuyo valor se sitúa entre 0 y 1, modulando la importancia que tienen las observaciones pasadas sobre el presente.

❖ Alisado doble de Holt

Se utiliza cuando la serie presenta tendencia, este método supone que la tendencia es lineal, pero su pendiente va variando en el tiempo. Se define como $x_t = L_t + b_t t + z_t$. donde L es la constante, b representa la pendiente, t el tiempo y z el ruido o componente irregular.

Este método depende del parámetro α visto en el alisado simple y de un nuevo parámetro β , que influye en la pendiente, modulando la importancia de las observaciones pasadas sobre la pendiente estimada en el tiempo t . Ambos parámetros oscilan entre 0 y 1. Así, para un valor de β cercano a 0, la pendiente será casi constante, mientras que, si es cercano a 1, la predicción de la pendiente se adapta al último valor observado dando más importancia al presente.

❖ Suavizado de Holt Winters

Se utiliza cuando la serie presenta estacionalidad. Si la incidencia de la estacionalidad no aumenta con el tiempo, el efecto debe modelarse con un modelo aditivo de esta forma: $X_t = (L_t + b_t) + S_t + z_t$. Por el contrario, si las variaciones estacionales aumentan con el tiempo debe modelarse con un modelo multiplicativo con esta forma: $X_t = (L_t + b_t) \cdot S_t + z_t$.

Como su nombre lo indica, el suavizado de Holt-Winters, parte del alisado de Holt, sin embargo, en este se aplica un triple alisado, además de los parámetros α visto en el alisado

simple y del parámetro β del alisado de Holt, Winters añade un tercer coeficiente γ . Este nuevo parámetro influye directamente sobre la estacionalidad.

Este nuevo parámetro γ toma valores entre 0 y 1, si es cercano a 0 la predicción en el tiempo t va a tomar un valor constante, dependiendo así de todas las observaciones pasadas dentro de ese mismo periodo, en cambio, si el valor se acerca a 1, la predicción depende solo de la observación hecha en el tiempo $t-p$, siendo p la frecuencia. Al ser este método de suavizado el que se aplicará en este trabajo, interesa conocer la nomenclatura con la que se obtiene la serie suavizada aplicando el modelo multiplicativo:

$$L_t = \alpha \frac{x_t}{S_{t-s}} + (1 - \alpha) (L_{t-1} + b_{t-1})$$

$$b_t = \beta (L_t - L_{t-1}) + (1 - \beta) b_{t-1}$$

$$S_t = \gamma \frac{x_t}{L_t} + (1 - \gamma) S_{t-s}$$

Donde L_t representa la constante, b_t es la pendiente y S_t es la estacionalidad. A través de este modelo, el cálculo de la predicción se define algebraicamente mediante la siguiente ecuación:

$$\hat{x}_{t+1} = (L_t + b_t) S_{t-s+1}$$

4.2.3 Modelos estacionarios

Los fenómenos dinámicos que observamos en series temporales pueden clasificarse como estacionarios cuando toman valores estables en el tiempo alrededor de un valor central, sin mostrar una tendencia o crecer o decrecer a lo largo del tiempo. Un proceso estocástico puede ser **estacionario en sentido estricto** o **estacionario en sentido débil**.

Una propiedad importante de los procesos estacionarios es que estos son estables ante combinaciones lineales. Un proceso estacionario muy simple que nos permitirá entender mejor lo dicho, es el del ruido blanco. Diremos que un modelo está ajustado perfectamente cuando el error resultante es un ruido blanco, es decir, cuando cumple estas condiciones:

- ❖ Media es igual a 0, $E[e_t] = 0, t = 1, 2, \dots$
- ❖ Varianza es constante, $Var(e_t) = \sigma^2, t = 1, 2, \dots$
- ❖ Las variables están incorrelados para todos los retardos, $Cov(e_t, e_{t-k}) = 0, k = \pm 1, \pm 2, \dots$

Los modelos estacionarios se clasifican de esta manera:

- ❖ **Modelos autorregresivos AR(p):** Se puede definir como modelos de regresión consigo mismo. Lo que hace es generalizar la idea de regresión para representar la relación entre una variable de la serie y las anteriores, es decir, se impone una dependencia temporal entre las variables del proceso.
- ❖ **Modelos de medias móviles MA(q):** Estos procesos son útiles para representar series de memoria corta, su identificación es sencilla ya que su función de autocorrelación se corta a partir de un determinado retardo
- ❖ **Modelo mixto ARMA (p,q):** Estos modelos juntan los dos anteriormente descritos, incluyen términos autorregresivos y términos de medias móviles.

4.2.4 Modelos no estacionarios

Los modelos no estacionarios, son aquellos que pueden mostrar tendencia, estacionalidad y otros efectos evolutivos en el tiempo. Un modelo puede ser no estacionario tanto en su media, en la varianza, en las autocorrelaciones o en otras características de la distribución. En la realidad, la mayoría de las series tienen un comportamiento no estacionario, por este motivo, interesan aquellos que se pueden convertir fácilmente en procesos estacionarios, por ejemplo, los procesos integrados al diferenciarlos.

Diremos que un proceso es integrado de orden 1 si la serie de las primeras diferencias $\omega_t = \nabla X_t = X_t - X_{t-1}$ ya es estacionaria. Si esta serie todavía no lo es, se seguirá diferenciando $\nabla \omega_t = \omega_t - \omega_{t-1} = X_t - 2X_{t-1} + X_{t-2} = \nabla^2 X_t$. En general, si se requieren hacer d diferencias, se dice que es un proceso integrado de orden d .

❖ **Modelo ARIMA (p,d,q)**

Se dice que un proceso es ARIMA (p,d,q) si al tomar diferencias del orden d , se llega a un proceso estacionario ARMA (p,q). La p representa el orden la parte autorregresiva estacionaria, la d , el orden de integración y la q la parte de media móvil. Este tipo de modelos se expresan de la siguiente forma:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) Z_t$$

Un modelo ARIMA (p,d,q) se caracterizan por tener un FAS con coeficientes positivos que decrecen de forma lineal, y que pueden ser distintos de cero hasta para valores altos de retardo.

❖ Modelo ARIMA Estacional (p,d,q)(P,D,Q)_s

Un modelo ARIMA es estacional cuando los datos tienen oscilaciones periódicas, al repetirse una cierta pauta cada s periodos. En este caso, se puede convertir una serie con estacionalidad en estacionaria mediante las diferencias del orden s , siendo s los periodos.

Si queremos diferenciar la serie estacionalmente al orden 1 o de periodo s , decimos que una diferencia estacional seria la diferencia entre una observación y la anterior de su mismo periodo. Al tomar diferencias de observaciones separadas s periodos, el proceso transformado es estacionario. En general, se define el operador diferencia estacional de orden D , de la siguiente manera:

$$\nabla_s^D X_t = (1 - B^s)^D X_t$$

Un modelo estacional general será por lo tanto de la forma ARIMA (p,d,q)(P,D,Q)_s donde los parámetros incluidos en el primer paréntesis corresponden a la parte no estacional y los del segundo a la parte estacional. Su ecuación general puede expresarse en términos del operador diferencial de la siguiente forma:

$$\left(1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{ps}\right) \left(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p\right) (1 - B^s)^D (1 - B)^d X_t = \left(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q\right) Z_t$$

4.3. Metodología Box Jenkins

La metodología Box Jenkins, denominada así en honor a los apellidos de los estadísticos George Box y Gwilym Jenkins, demuestra cómo se pueden ajustar a series reales los modelos ARMA o ARIMA, a través de cuatro etapas explicadas a continuación:

1. Identificación

En esta fase se utilizan los datos históricos de la serie para identificar el posible modelo ARIMA que sigue la serie. En esta primera etapa se debe:

- ❖ Decidir qué transformaciones aplicar para convertir la serie observada en estacionaria, obteniendo una media y varianza constantes.
- ❖ Determinar un modelo ARMA para la serie estacionaria, es decir, los órdenes p y q de su estructura autorregresiva y de media móvil.

2. Estimación

Esta segunda etapa se estima los parámetros AR y MA, usando el método de mínimos cuadrados o máxima verosimilitud, se obtienen los errores estándar y los residuos del modelo. El objetivo, será obtener los parámetros del modelo $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2, u$ y seleccionar el modelo ARMA de entre los estimados.

3. Diagnósis

Una vez obtenidos los residuos, se comprueba que estos no tengan estructura de dependencia y por lo tanto sigan un proceso de ruido blanco. Como método de diagnóstico simple y habitual, se suele dibujar dos líneas paralelas a $\frac{2}{\sqrt{T}}$ distancia del origen tanto del gráfico del FAS y FAP estimadas y comprobar que los coeficientes estén dentro de dichas bandas que representan los intervalos de confianza para las autocorrelaciones cero.

❖ Medidas de adecuación del modelo:

Una vez aceptado el modelo, podemos compararlo con otros de orden superior calculando los errores ε cometidos con cada modelo para las T observaciones de las que disponemos. $\varepsilon_t = X_t - \hat{X}_t$. El valor total de estos residuos se resume en distintos estadísticos.

El primer estadístico que permitirá comparar los modelos y que a partir de medidas absolutas permite definir las medidas en término relativo es el R^2 . Expresado de esta forma:

$$R^2 = 1 - \frac{\sum_{t=1}^T SSE}{\sum_{t=1}^T SST}$$

En segundo lugar, se utilizan también dos medidas estadísticas importantes como son: el **Criterio de información de Akaike (AIC)** y el **criterio bayesiano de Schwarz (SBC o BIC)**. Ambos basados en el logaritmo de la función de verosimilitud utilizada para calcular los estimadores de la serie, para utilizarlos debemos asegurarnos de que los residuos tengan una distribución normal. El AIC es el que realmente mide el desajuste, mientras que el BIC que incluye penalización, mide la complejidad del modelo a partir del número de parámetros.

$$AIC = -2\ln(L) + 2k$$
$$BIC = -2\ln(L) + \ln(n)k$$

Donde L es la función de verosimilitud, k el número de parámetros y n el número de residuos calculados. Mientras menor sea su valor, mejor será el modelo.

4. Predicción

Después de que el modelo ha sido construido, lo usamos para realizar predicciones de valores futuros de la variable estudiada. Para las predicciones se debe tomar en cuenta que los parámetros de las funciones y residuos $\varepsilon_1, \varepsilon_2 \dots$ presentes y pasados son conocidos. De igual manera resulta interesante comparar la predicción con los datos reales y así observar lo acertado que ha resultado el modelo elegido.

En la siguiente figura podemos tener una clara idea del proceso de aplicación de un modelo ARIMA, donde se refleja la aplicación de la metodología de Box-Jenkins.

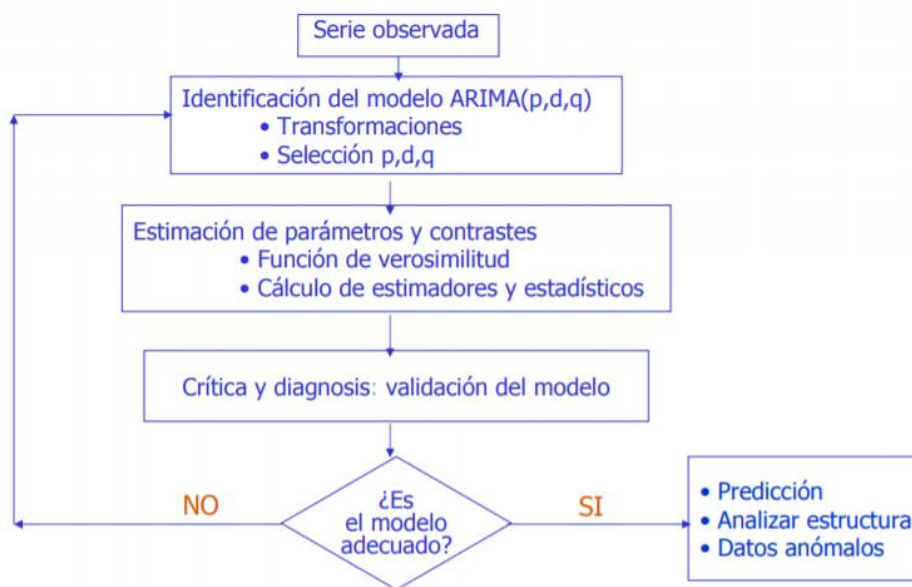


Figura 2: Proceso de adecuación de modelos ARIMA

4.4 Modelos de regresión lineal

La regresión lineal permite generar un modelo en el que el valor de una o más variables dependientes (Y) se determina a partir de un conjunto de variables independientes llamadas predictores (X1, X2, X3...). Se define de esta forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Donde β_0 son los coeficientes parciales de regresión y ε es el error.

Los modelos de regresión múltiple pueden emplearse para predecir el valor de la variable dependiente o para evaluar la influencia que tienen los predictores sobre ella, lo segundo es lo que se buscará en este trabajo. La magnitud de cada coeficiente parcial de regresión depende de las unidades en las que se mide la variable predictora a la que corresponde, por lo que no está asociada con la importancia de cada predictor. Para poder determinar qué impacto tienen en el modelo cada una de las variables, se emplean los **coeficientes parciales estandarizados**, que se obtienen al restar su media y dividir entre la desviación estándar.

Idealmente, en los modelos de regresión lineal múltiple las variables predictoras deben ser independientes, es decir, no debe haber multicolinealidad entre ellos. **La multicolinealidad** ocurre cuando un predictor está linealmente relacionado con uno o varios de los otros predictores del modelo o cuando es la combinación lineal de otros predictores. Existen varios métodos para corregir la multicolinealidad, siendo dos de ellos los que veremos a continuación.

4.4.1 Regresión LASSO

La Regresión LASSO (Least Absolute Shrinkage and Selection Operator) es un modelo de regresión lineal para datos de alta dimensión y fue propuesta por Tibshirani[11]. Es una técnica de regresión lineal regularizada que, mediante la contracción de los coeficientes, logra estabilizar las estimaciones y predicciones y que puede verse como un método de selección de variables. La dificultad consiste en determinar el **valor óptimo de t/λ** .

Los estimadores se obtienen al minimizar la suma de cuadrados penalizada para cada landa:

$$\min \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^k |\beta_j|$$

A partir de cierto valor del parámetro de penalización (λ) el estimador de Lasso produce estimaciones nulas para algunos coeficientes y no nulas para otros, con lo cual Lasso realiza una especie de selección de variables en forma continua, debido a las propiedades del valor absoluto.

4.4.2 Regresión PLS

La Regresión PLS o de Mínimos Cuadrados Parciales es un modelo de regresión lineal para datos con Multicolinealidad. Lo que hace es transformar las variables explicativas en componentes ortogonales, dando solución al problema de multicolinealidad y haciendo una reducción de la dimensionalidad del espacio de variables predictoras. Este algoritmo reduce el número de predictores mediante el uso de una técnica similar a las del análisis de componentes principales para extraer un conjunto de componentes que describa la correlación máxima entre los predictores y las variables de respuesta. La regresión PLS se diferencia entre una PLS1 aplicada cuando el modelo tiene una única variable respuesta Y o PLS2, aplicada cuando existe más de una variable respuesta.

En este trabajo se aplicará justamente una regresión PLS2, por lo que teniendo un conjunto de variables a explicar $Y = (y_1, y_2, \dots, y_{n \times q})$ que tratamos de relacionar con otro conjunto de variables explicativas o predictoras $X = (x_1, x_2, \dots, x_{n \times p})$ el algoritmo de regresión PLS2 efectúa una reducción de la dimensionalidad de un conjunto de variables X , bajo la condición de que estas componentes principales sean también las más explicativas posibles respecto del conjunto de variables Y . En este caso es posible predecir las variables y_k a partir de las x_j separando mejor lo que es común a los datos de aquello que es más específico. [13]

Capítulo 5

5. Desarrollo del Análisis

Como se ha mencionado desde un principio, es importante recordar que, para el desarrollo de este trabajo se han usado ficheros de datos distintos que permitirán tener una idea clara del contexto general de estudiantes en España. Además, centraremos nuestro análisis específicamente en los estudiantes extranjeros extracomunitarios, aquellos que no tienen nacionalidad o residencia de un país que forme parte de la Unión Europea y que requieren de una autorización de estancia por estudios para movilizarse a España. Al observar la figura 3 tendremos una idea clara del universo de nuestros datos y el proceso para llegar a los que más nos interesan.

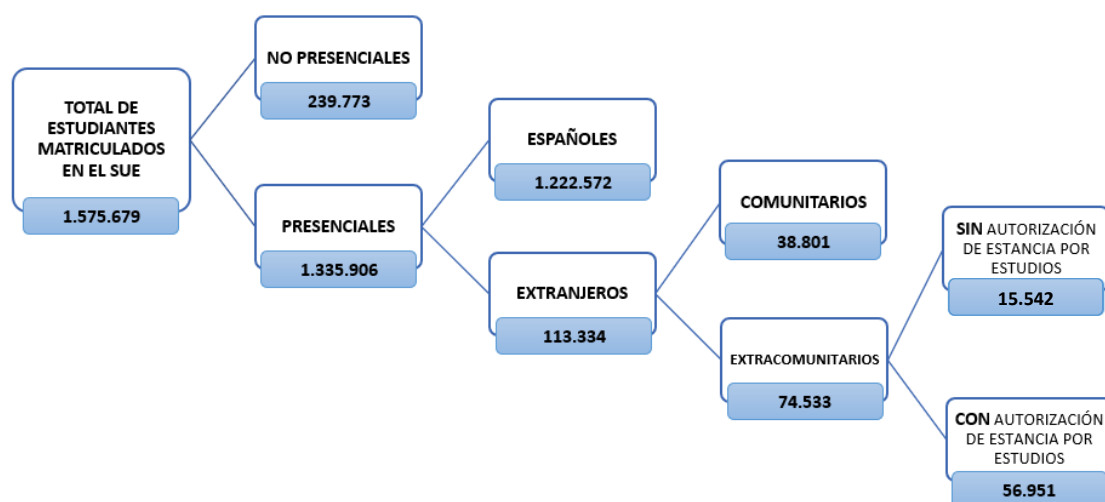


Figura 3: Universo y desagregación de datos utilizados

Se utilizarán datos de los últimos 25 años en el estudio, desde 1995 hasta 2018 y/o 2019 en caso de una actualización de los mismos en ciertas fuentes. Como conocimiento previo, decimos que España en el año 2018, registró un total de estudiantes extranjeros en modalidad presencial de 113.334⁵, de los cuales 56.951 corresponden a estudiantes extracomunitarios que han solicitado una autorización de estancia por estudios. En los últimos 25 años el aumento de estudiantes extranjeros ha tenido en general una tendencia positiva, en promedio un aumento anual del 9% tomando en cuenta todos estos años.

⁵ Ministerio de Ciencia, Innovación y Universidades (2018). Estadística de Internacionalización. Movilidad de estudiantes extranjeros.

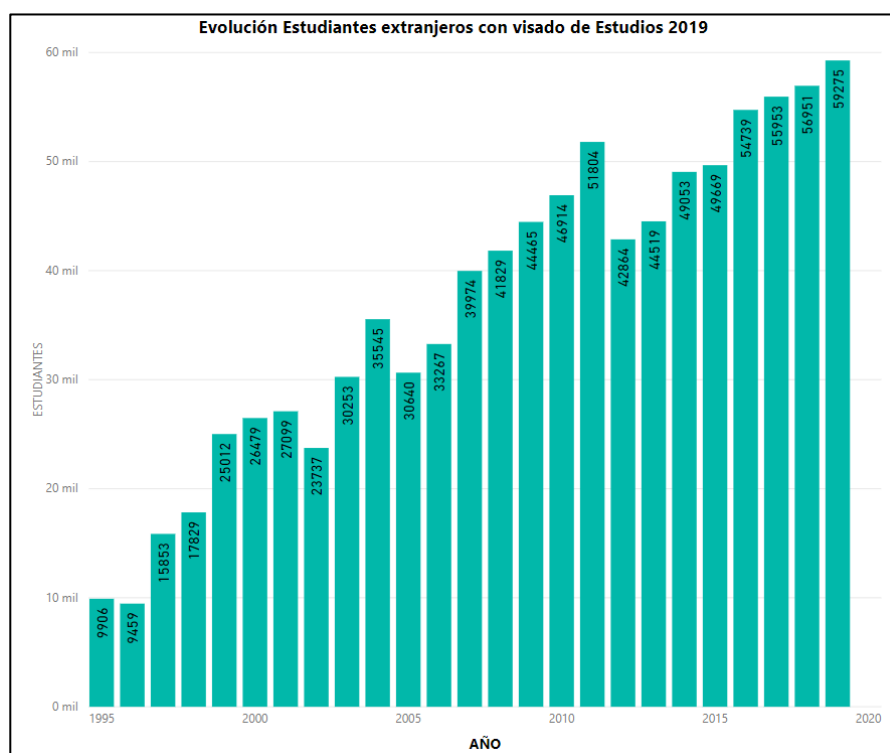


Figura 4: Evolución de estudiantes extranjeros desde 1995

Vale la pena recalcar que hay años en donde ha habido una disminución de estudiantes extranjeros, la más importante entre el año 2011 y 2012, donde hubo una reducción del 17,26%. En contraste a esto, el año con mayor aumento en la variación de estudiantes fue entre 1996 y 1997, donde se dio un aumento del 67,60 % de personas que obtuvieron su visa por estancia por estudios en España. Si tomamos en cuenta los últimos tres años, el aumento ha sido de 2,7% en promedio cada año, con una cifra de estudiantes por año bastante alta, superando en todos los casos 55.000 estudiantes que han solicitado la estancia por estudios.

5.1. Análisis y descripción de las Variables

En este apartado se busca describir las variables a utilizarse en el desarrollo de este trabajo, se mostrará la distribución, evolución y datos más importantes de las principales variables relativas a los estudiantes extranjeros en España. Este análisis inicial, servirá también para una comprensión del contexto general y del perfil de los estudiantes extranjeros.

5.1.1 Sexo

Esta variable muestra información sobre el sexo de los estudiantes extranjeros. En España, los estudiantes extranjeros han ido aumentando continuamente, llama la atención que este aumento siempre ha ido de la mano de una mayor cantidad de mujeres que obtienen su visado de estudios, para esta comparación se han tomado en cuenta los últimos 17 años, teniendo así, a partir del 2002, datos más exactos.

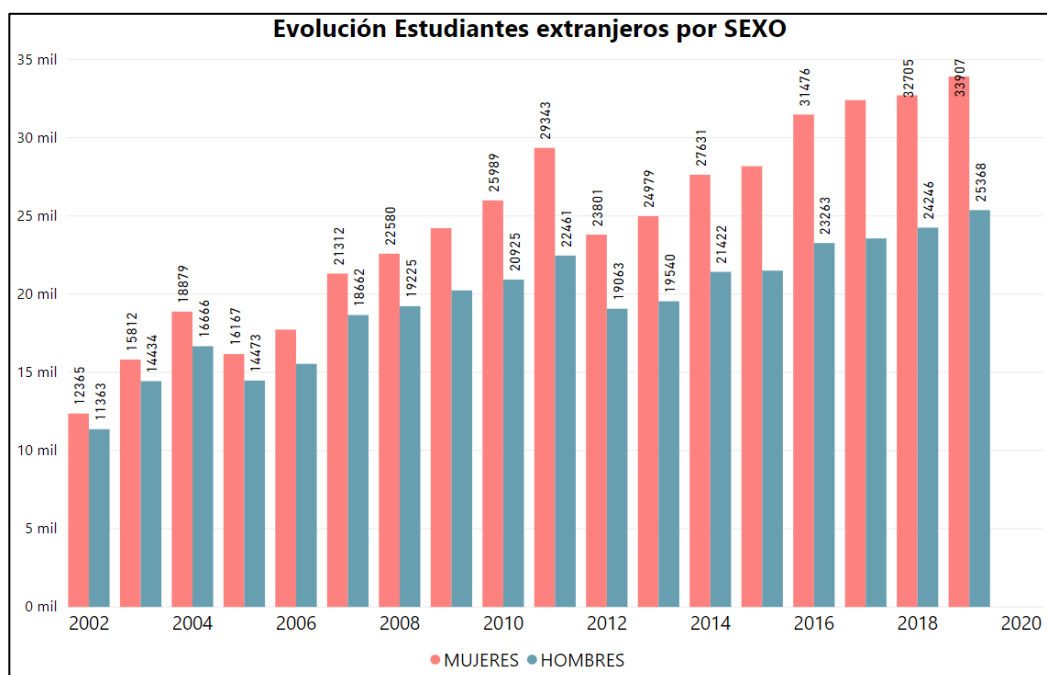


Figura 5: Evolución y distribución de Extranjeros por Sexo

En todos los años hay más mujeres que hombres, en promedio tomando en cuenta todos estos años, hay un 10% más de mujeres que solicitan autorización de estancia por estudios. En el año 2019, hay 33.907 mujeres frente a 25.368, exactamente un 14,41% de diferencia, por lo que afirmamos que se sigue una tendencia de aumento de mujeres frente a hombres.

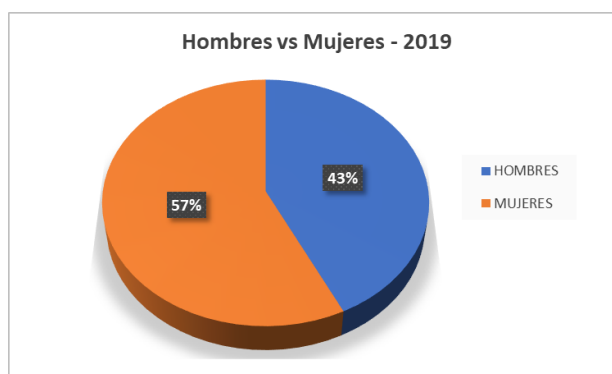


Figura 6: Distribución de estudiantes extranjeros en España por Sexo

5.1.2. Edad

La segunda variable que se analizará es la edad de los estudiantes extranjeros en España. Para el análisis, se usarán seis grupos distintos de rangos de edad; de 18 a 24 años; de 25 a 29 años; de 30 a 34 años; de 35 a 39 años; de 40 años o más.

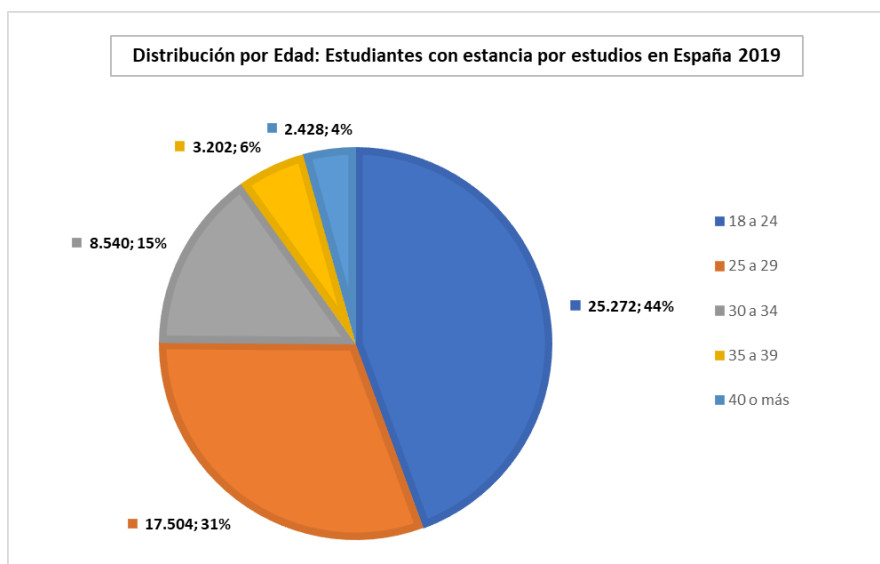


Figura 7: Distribución de estudiantes extranjeros por Edad

Como es de esperarse, la mayor cantidad de estudiantes se encuentran en el rango de edad entre los 18 y 24 años, representando el 44% del total y siendo principalmente estudiantes que vienen a España a realizar sus estudios de grado. En segundo lugar, están los estudiantes del siguiente rango de edad, de entre 25 y 29 años representando el 31% y el siguiente grupo de entre 30 y 34 que representan el 15%, estos dos últimos los podemos relacionar más con estudios de posgrado y algunos de doctorado. Finalmente, están los grupos de edad de 35 a 39 años y de 40 en adelante que representan un 6 y 4%, respectivamente.

5.1.3. Región de procedencia

A partir de los datos recopilados, se establecen seis regiones principales de procedencia de los estudiantes extranjeros que son: el Resto de Europa, África, América del Norte, América Central y del Sur, Asia y Oceanía. Tomando de ejemplo el año 2019, las personas extranjeras con estancia por estudios procedentes de América Central y Suramérica representan el 40,6% del total. En segundo lugar y debido en gran medida a los convenios de España con China, vemos una gran cantidad de estudiantes procedentes de Asia, representando el 26,4% del total.

América del Norte es otra de las grandes potencias de exportación de estudiantes a todo el mundo, en este caso representan en España el 17% del total de estudiantes extranjeros en 2019. Para finalizar, África viene a ser la región que ocupa el quinto lugar, con un porcentaje de casi el 10% de estudiantes, el Resto de Europa (países fuera de la UE) representan el 5,76% y finalmente Oceanía representa tan solo el 0,49% respecto al total.

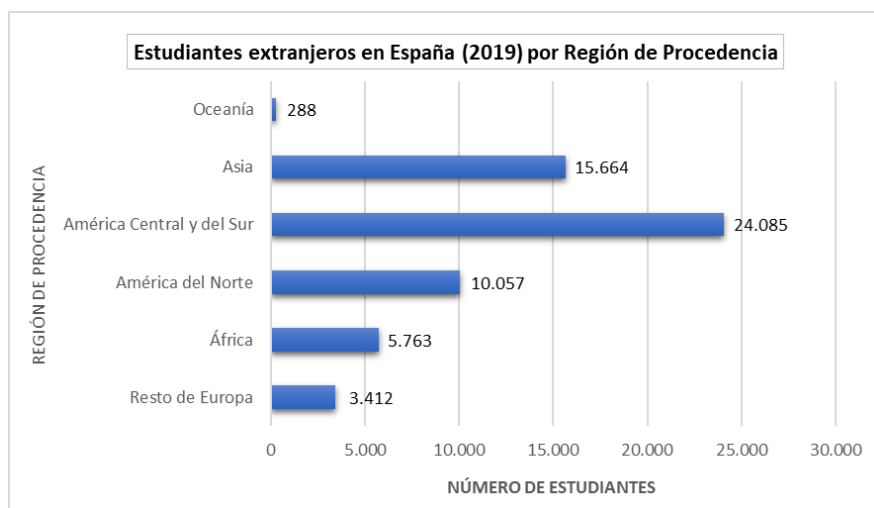


Figura 8: Estudiantes extranjeros en España por Región de procedencia

Si observamos la evolución de estos datos en los últimos años vemos el claro despunte de Asia, prácticamente duplicando su número de estudiantes en España con un 97,6% de aumento en 10 años. En el caso de América Central y del Sur, se ha mantenido siempre la hegemonía de esta región, aunque entre 2011 y 2013 hubo un declive que se puede asociar a la crisis de esos años, a partir de allí ha habido un continuo crecimiento con un aumento promedio del 4% tomando en cuenta los últimos 5 años.

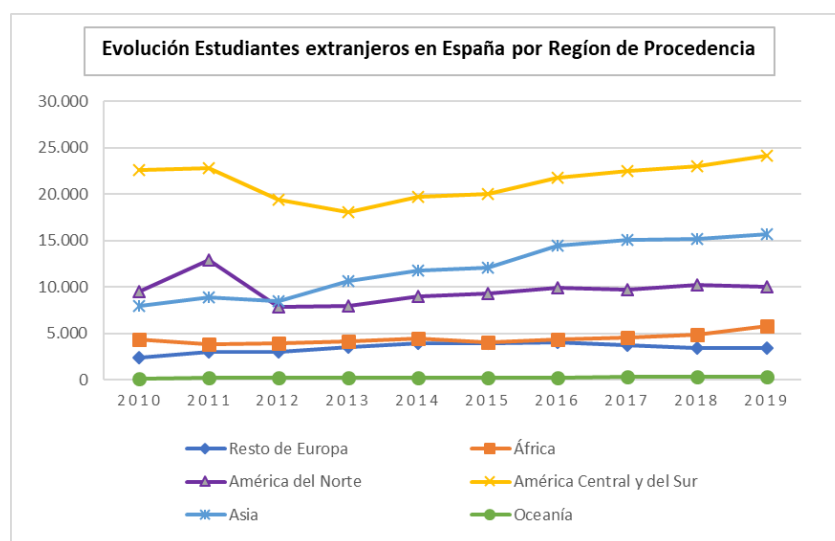


Figura 9: Evolución de estudiantes por Región de procedencia desde 2010

5.1.4. País de procedencia

Continuando con la misma línea relativa a la procedencia de los extranjeros con estancia por estudios en España, resulta interesante analizar las principales nacionalidades o países de donde llegan la mayoría de los estudiantes, para ello haremos un ranking de los diez países que movilizaron más estudiantes hacia España en el 2019.

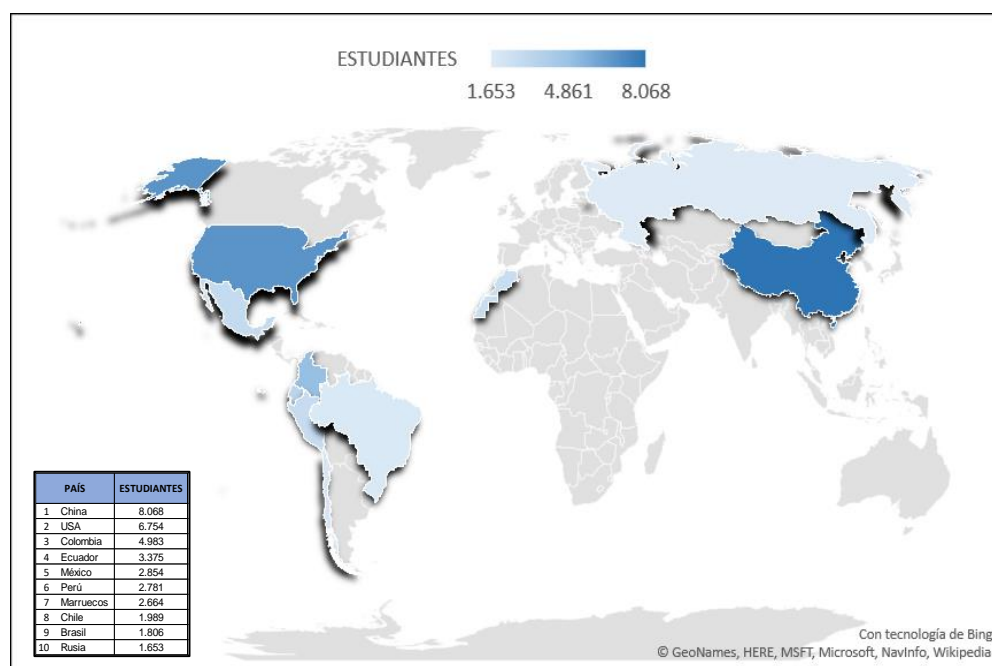


Figura 10: País de procedencia de la mayor cantidad de estudiantes extranjeros

En las siguientes tablas vemos que China y Estados Unidos ocupan el primer lugar, a este último se suma México de la misma región, luego figuran 5 países de América del Sur, Marruecos de África y solo Rusia del Resto de Europa. Es interesante comparar la cantidad de estudiantes de cada país, en relación a su población total, donde el ranking cambia notoriamente, China pasa a ser el número diez, mientras que Ecuador y Colombia pasan a liderar la tabla, demostrando así la importancia de la región Latino Americana.

PAÍS	ESTUDIANTES	% del Total
1 China	8.068	13,61%
2 USA	6.754	11,39%
3 Colombia	4.983	8,41%
4 Ecuador	3.375	5,69%
5 México	2.854	4,81%
6 Perú	2.781	4,69%
7 Marruecos	2.664	4,49%
8 Chile	1.989	3,36%
9 Brasil	1.806	3,05%
10 Rusia	1.653	2,79%

PAÍS	ESTUDIANTES	POBLACIÓN TOTAL
1 Ecuador	3.375	17.300.000
2 Colombia	4.983	48.258.494
3 Chile	1.989	19.107.216
4 Perú	2.781	32.495.510
5 Marruecos	2.664	35.330.305
6 México	2.854	126.577.691
7 Marruecos	2.664	35.330.305
8 Rusia	1.653	146.804.372
9 Brasil	1.806	210.385.000
10 China	8.068	1.403.500.365

Tabla 3: Ranking de países con más estudiantes extranjeros en España

5.1.5. Comunidad Autónoma de destino

Una vez analizado el origen de los estudiantes extranjeros, mediante esta variable se estudia la Comunidad Autónoma (CCAA) de destino donde realizan sus estudios. En el caso de España, la CCAA a la que se movilizan la mayoría de los estudiantes extranjeros es Cataluña, que recibe a un 28.53% del total de estudiantes. Le sigue muy de cerca Madrid, acogiendo a 28.27%. Es importante mencionar también a Ceuta y Melilla que no se muestran en el mapa y entre las dos suman un total de 34 estudiantes extranjeros en el 2019. En la siguiente figura observamos los datos exactos y la proporción de estudiantes según cada CCAA, siendo las más oscuras las que mayor cantidad de estudiantes reciben sobre el total y las más claras las que menos.

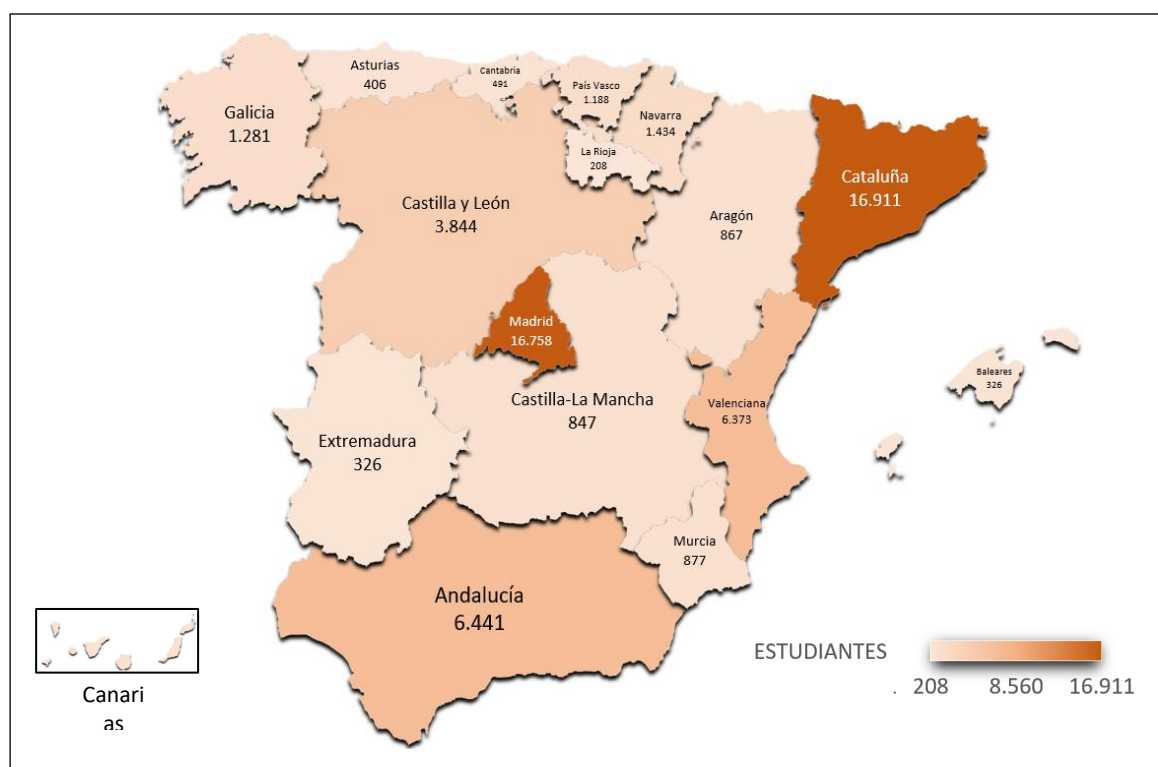


Figura 11: Distribución de estudiantes extranjeros por CCAA en 2019

Como es de esperarse, las Comunidades Autónomas donde se encuentran las principales ciudades de España y donde hay una mayor y mejor calidad de oferta educativa son las que más cantidad de estudiantes extranjeros reciben. Por este motivo, resulta también interesante mostrar una lista de las principales ciudades en donde se solicitan más autorizaciones de estancia por estudios, para la elección de estas se han tomado en cuenta aquellas que han recibido más del 2% de estudiantes con autorización de estancia por estudios en el último año.



Figura 12: Ciudades con más del 2% del total de estudiantes extranjeros

La ciudad que más estudiantes con autorización de estancia por estudios recibe, es Madrid, seguida por Barcelona. Entre las dos representan casi el 53% del total, luego está Valencia, una ciudad que recibe muchos estudiantes tomando en cuenta su tamaño, representando el 8,19% del total y que dentro de su CCAA se apoya por Alicante que aparece también en este listado. Las ciudades principales del sur de España acogen también muchos estudiantes universitarios, dentro del ranking aparece Granada, Málaga y Sevilla. En el norte, Pamplona viene a ser la ciudad elegida por los estudiantes extranjeros y también al hablar de una comunidad autónoma distinta, Salamanca es la ciudad que representa a Castilla y León y que recibe anualmente un 3,46% del total de estudiantes extranjeros en España. Para analizar mejor esta variable, vemos su evolución en los últimos diez años.

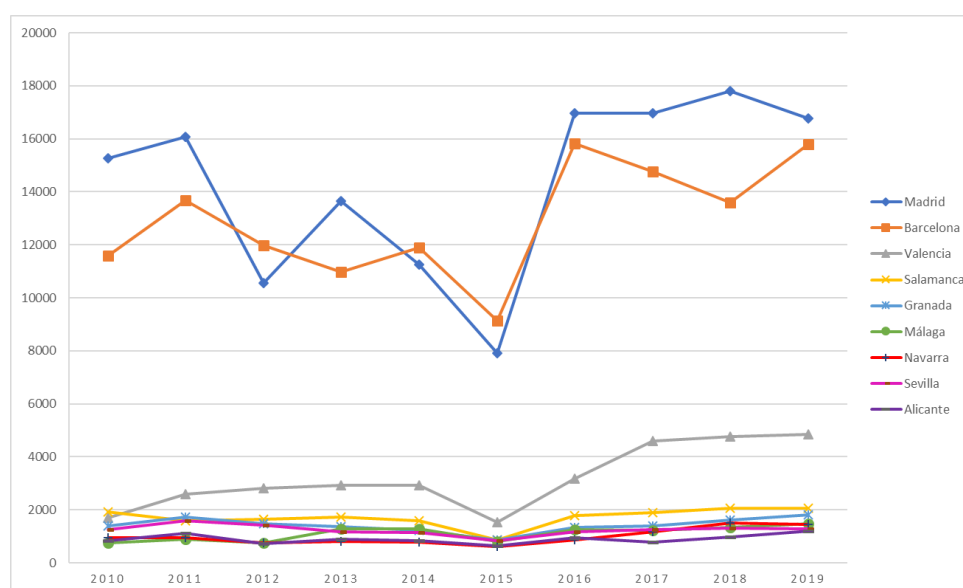


Figura 13: Evolución de extranjeros con estancia por estudios por ciudad de destino.

Distinguimos cómo ha cambiado la llegada de estudiantes extranjeros a las principales ciudades, en general se observa que Madrid ha mantenido la hegemonía en la mayoría de los años, sin embargo, en algunos de esos años, Barcelona ocupa el primer lugar. Fijándonos en el resto de las ciudades, Valencia destaca sobre el resto y en los últimos 5 años ha tenido una tendencia muy positiva, recibiendo cada año más estudiantes extranjeros. Las demás ciudades tienen un comportamiento similar con un número de estudiantes anuales que se mantiene en aumento desde el 2015, como se puede observar en la siguiente figura.

5.1.6. Nivel de Estudios

Como se mencionó en el Capítulo 2, en esta investigación nos centraremos en las personas extranjeras que realizan sus estudios terciarios en España, entre los que se distinguen tres niveles: Grado, Máster o Posgrado y Doctorado. Para el análisis de esta variable, se utilizarán datos procedentes del Sistema Universitario Español (SUE) del año 2018 obtenidos a través del Ministerio de Educación y Formación Profesional, donde se toma en cuenta a todos los estudiantes extranjeros matriculados en el SUE sin importar si requieren o no autorización de estancia por estudios, por lo que el Universo de datos es superior al usado en el análisis de las anteriores variables.

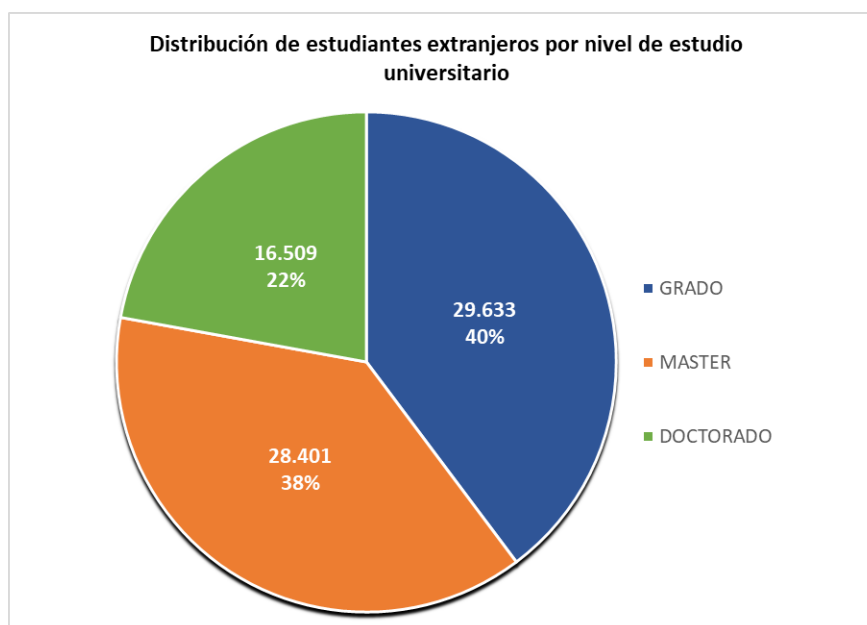


Figura 14: Distribución de estudiantes extranjeros por Nivel de estudios

En la figura anterior, se toma en cuenta un total de 74.543 estudiantes extranjeros que se han matriculado tanto en estudios de Grado, Master o Doctorado en España en el año 2018. Del total de los estudiantes extranjeros, un 40% se matriculan en estudios de grado, dato que va de la mano y se relaciona con el rango de edad de entre 18 y 24 años al que pertenecen la mayoría de los estudiantes. En segundo lugar, existe una gran demanda de estudios de Máster, en los que se matriculan el 38% del total de estudiantes extranjeros, finalmente, el 22% de estudiantes matriculados en el sistema universitario optan por estudios de Doctorado, mismos que tienen como prerequisite indispensable, los dos estudios mencionados anteriormente.

Con relación a los resultados al analizar la variable sexo, donde se distinguió que hay un mayor número de estudiantes extranjeros que son mujeres; a continuación, se muestra que, en estudios de grado, hay una clara diferencia entre el sexo de los estudiantes, habiendo un 18% más mujeres que hombres. En estudios de Máster, esta diferencia disminuye casi a la mitad, al estar matriculadas 54% de mujeres frente a un 46% de hombres. Finalmente, en el nivel de estudio de doctorado, las cosas cambian, pues en total hay más hombres, con una proporción del 53%, mientras que las mujeres representan un 47% del total de extranjeros matriculados.

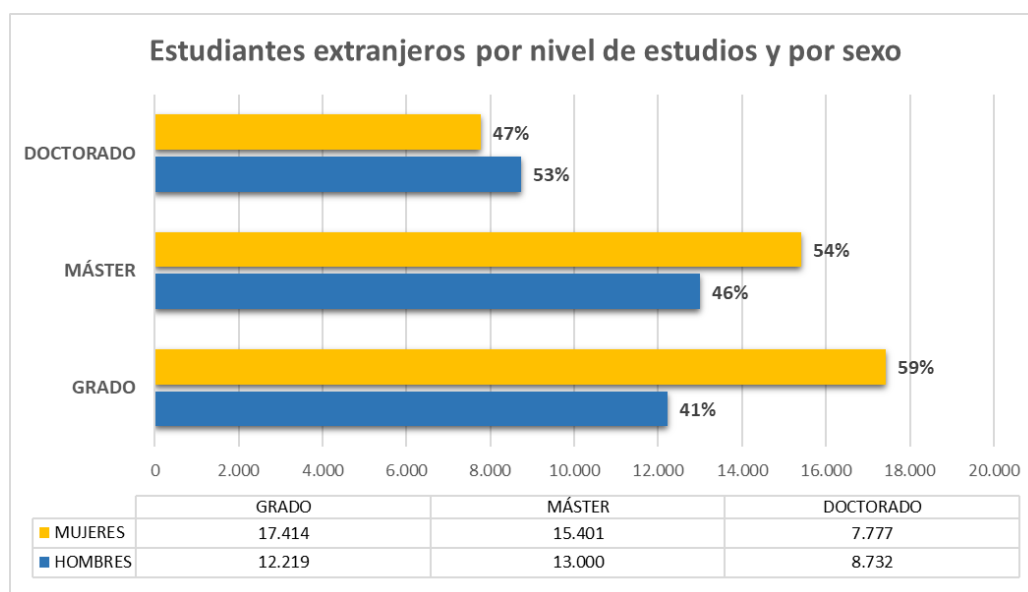


Figura 15: Estudiantes extranjeros por Nivel de estudios y por sexo

- **Relación entre Nivel de estudios y la Comunidad Autónoma de destino**

Dentro del análisis de esta variable, es interesante establecer una relación con las CCAA de destino. Como era de esperarse, Madrid y Cataluña son las Comunidades con mayor número de estudiantes en los tres niveles de estudio, ambas tienen más estudiantes en nivel de Máster, luego de grado y finalmente doctorado. En tercer lugar, la Comunidad Valenciana tiene una distribución normal, tiene más estudiantes de grado, luego de Máster y menos de doctorado. Andalucía en cambio, tiene mayoría de estudiantes de grado, pero luego de doctorado, lo que llama la atención, pues al final están los de Máster. Por último, resalta Castilla y León, donde sus datos se asemejan a los de Madrid y Cataluña. Finalmente vemos que el resto de CCAA tienen distribuciones similares y con más estudios a nivel de grado, luego máster y doctorado, destacando en último lugar La Rioja con tan solo 140 estudiantes extranjeros matriculados en sus distintos niveles de estudios.

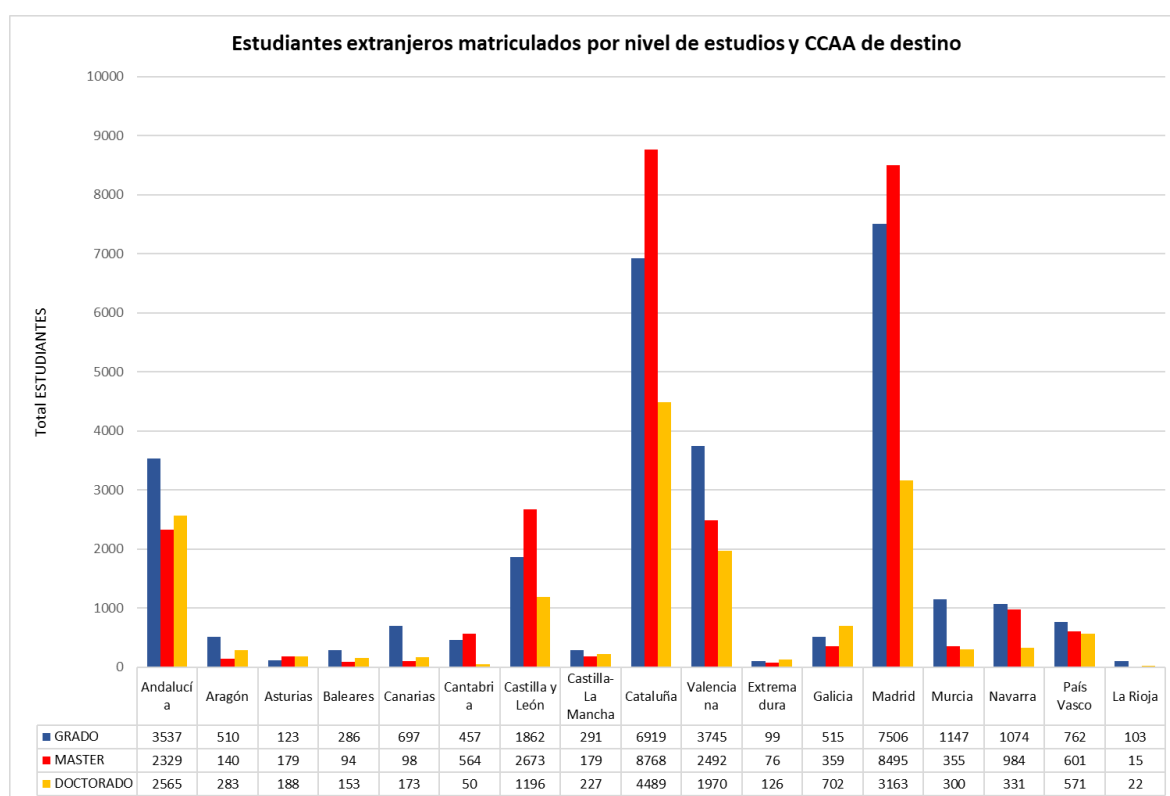


Figura 16: Estudiantes extranjeros por Nivel de estudios y CCAA de destino – 2018.

5.1.7. Tipo de Institución Universitaria

Este análisis no estaría completo sin hacer referencia a la distribución de los estudiantes extranjeros por universidades, al ser estas las protagonistas de la captación de los estudiantes. Existen tres tipos de instituciones de educación superior donde se imparten enseñanzas universitarias, Universidades Públicas, Universidades Privadas y centros universitarios privados⁶. En este análisis, usando datos del Sistema Universitario Español (SUE) buscaremos distinguir cómo se distribuyen los estudiantes extranjeros según la institución universitaria donde deciden realizar sus estudios. Esto permitirá tener una idea clara de las preferencias y decisiones de las personas extranjeras que optan cursar sus estudios terciarios en España. Para tener información más precisa, se tomarán en cuenta únicamente los programas presenciales universitarios.

Existe una clara preferencia por Universidades Públicas frente a Privadas en el caso de estudiantes extranjeros matriculados, lo que puede deberse principalmente a la mayor oferta del sector público, el buen reconocimiento académico de estas instituciones y a un tercer factor muy importante que son los precios, que son muy inferiores frente a los estudios en instituciones privadas. Observamos en la figura 18, que el 82% de estudiantes extranjeros se matriculan en Universidades públicas frente a un 18% que lo hacen en Universidades privadas.

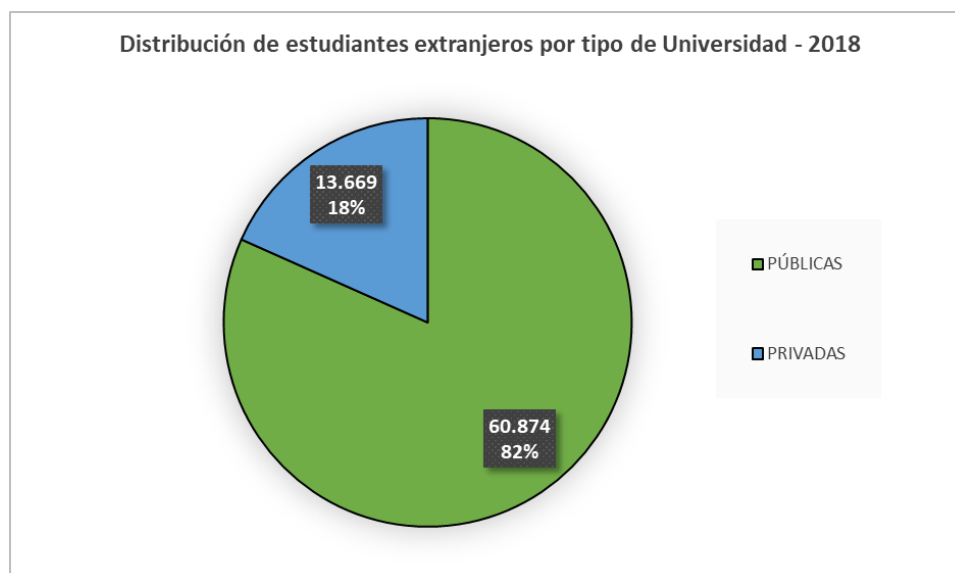


Figura 17: Distribución de estudiantes extranjeros por tipo de Universidad (2018).

⁶ Los centros universitarios privados deben estar integrados en una universidad privada como centros propios de la misma, o bien deben adscribirse a una universidad pública o privada.

Comparando el número de estudiantes extranjeros y su representación en el universo total de estudiantes por cada nivel, incluyendo a los estudiantes españoles y comunitarios. Decimos que en los estudios de máster y doctorado claramente es donde tienen mayor importancia la cifra de matriculados extranjeros, por este motivo, resulta interesante desagregar el nivel de estudios y profundizar más aún. Eligiendo el nivel de Máster como el más representativo, estudiaremos la distribución en las distintas universidades tanto de carácter público y privado de los extranjeros matriculados únicamente en estudios de este nivel, centrándonos en una muestra considerable de un total de 28.401 estudiantes extranjeros matriculados en Masters presenciales para el periodo 2018-2019.

	GRADO		MÁSTER		DOCTORADO	
ESTUDIANTES MATRICULADOS	TOTAL	%	TOTAL	%	TOTAL	%
TOTAL DE ESTUDIANTES	1.102.678	100,0%	160.449	100,0%	84.184	100,0%
Universidades Públicas	959.885	87,1%	124.066	77,3%	79.751	94,7%
Universidades Privadas	142.793	12,9%	36.383	22,7%	4.433	5,3%
TOTAL DE EXTRANJEROS	29.633	2,7%	28.401	17,7%	16.509	19,6%
Universidades Públicas	23.821	2,2%	21.270	13,3%	15.783	18,7%
Universidades Privadas	5.812	0,5%	7.131	4,4%	726	0,9%

Tabla 4: Extranjeros matriculados por tipo de centro y nivel de estudios

- **Distribución por Universidades**

Una vez diferenciados los estudiantes extranjeros matriculados en estudios de máster, podemos analizar la distribución de estos entre las principales universidades españolas, sean estas públicas o privadas. En la Tabla 6 se presenta un listado con las universidades más destacadas por la presencia de estudiantes internacionales. La lista se ha confeccionado ordenando de mayor a menor las universidades presenciales tanto por el número de estudiantes extranjeros como por su porcentaje respecto al total. Además, se ha tomado en cuenta aquellas Universidades que tienen una representación mayor al 1,5% del total y se ha verificado que la lista incluya las Universidades más importantes del país según el Ranking ARWU⁷ de Shanghai 2018.

⁷ Ranking Académico Mundial de Universidades 2018. Recuperado de <http://www.shanghairanking.com>.

TOP UNIVERSIDADES PÚBLICAS		
UNIVERSIDAD	ESTUDIANTES EXTRANJEROS	% del TOTAL
Rey Juan Carlos	2.541	10,7%
Universidad de Barcelona	2.393	10,0%
Politécnica de Catalunya	1.426	6,0%
Complutense de Madrid	1.380	5,8%
Autónoma de Barcelona	1.144	4,8%
Politécnica de València	1.138	4,8%
Universidad de València	1.014	4,3%
Pompeu Fabra	960	4,0%
Universidad de Granada	812	3,4%
Universidad de Sevilla	777	3,3%
Carlos III de Madrid	688	2,9%
Politécnica de Madrid	679	2,9%
Alcalá	592	2,5%
País Vasco	493	2,1%
Autónoma de Madrid	460	1,9%
Salamanca	425	1,8%
TOP UNIVERSIDADES PRIVADAS		
IE Universidad	1.589	27,3%
Ramón Llull	927	15,9%
Navarra	920	15,8%
Europea del Atlántico	859	14,8%
Antonio de Nebrija	687	11,8%
Europea de Madrid	682	11,7%
Católica San Antonio de Murcia	263	4,5%
Francisco de Vitoria	233	4,0%
Camilo José Cela	197	3,4%
Pontificia Comillas	141	2,4%
Vic-Central de Catalunya	134	2,3%
Deusto	126	2,2%
Internacional de Catalunya	93	1,6%

Tabla 5: Estudiantes extranjeros de Máster en principales universidades de España

Como se muestra en la tabla anterior, las universidades seleccionadas se acumulan principalmente en las comunidades con mayor proporción de estudiantes internacionales. Madrid, con nueve centros, acumula el mayor número de universidades. Hay que destacar, además, las primeras posiciones de un pequeño grupo de universidades privadas donde su porcentaje de extranjeros es muy alto, por un lado, y de las universidades públicas de la ciudad de Barcelona, por otro. En Madrid la universidad pública que acoge a un 10,7% de total de extranjeros es la Rey Juan Carlos, en Barcelona; representando el 10% del total aparece la Universidad de Barcelona.

En el grupo de las Universidades privadas, destaca el caso de la IE University de Madrid, con un porcentaje de extranjeros del 27%. En Barcelona, la Universidad Ramón Llull acumula casi el 16% del total de extranjeros y en tercer lugar, con un porcentaje de representación muy cercano al anterior (15,8%), está la Universidad de Navarra e Pamplona.

En la figura 18, vemos la distribución geográfica de estas Universidades.



Figura 18: Cantidad de Universidades por Comunidad Autónoma

5.1.8. Precios

La última variable que se analizará en este trabajo es la de precios, principalmente refiriéndonos al precio de matrícula por crédito que pagan los estudiantes extranjeros que realizan sus estudios en España y tomando en cuenta las variaciones existentes, en segundo lugar, se evaluará el precio de alquiler de una habitación en piso compartido, uno de los factores claves que influyen en los estudiantes para decidir el destino de sus estudios universitarios.

5.1.8.1 Precios de estudios Universitarios

Uno de los principales determinantes de la elección de una institución académica, sea pública o privada por parte de los estudiantes extranjeros es el precio de la matrícula. No solo existe una gran diferencia por el tipo de institución, sino también dentro de las instituciones varían los precios siendo estos más bajos para estudiantes españoles y comunitarios⁸ y mayores

⁸ Estudiantes extranjeros que tienen nacionalidad de un país miembro de la Unión Europea.

para estudiantes extranjeros o extracomunitarios. Los precios se establecen por crédito universitario y dependen del nivel de estudios, distinguiendo entre estudios de grado, estudios de máster habilitante, estudio de máster no habilitante⁹ y estudios de doctorado.

❖ Precios públicos para estudiantes españoles y comunitarios

Para tener una idea más clara de las diferencias existentes, es importante conocer los precios pagados por españoles y ciudadanos comunitarios. En la figura 19 se muestra esta información, el valor del crédito en primera matrícula en cada CCAA y en cada nivel de estudio. Podemos decir, usando un caso hipotético que, si un español desea estudiar un máster habilitante y elige Madrid como destino pagaría 30,84€ por crédito, pero si se matricula en Andalucía, pagaría 13,68€, es decir un 130% más en la capital, por esto resulta importante ver los precios en general y tener una idea clara de la variación de estos.

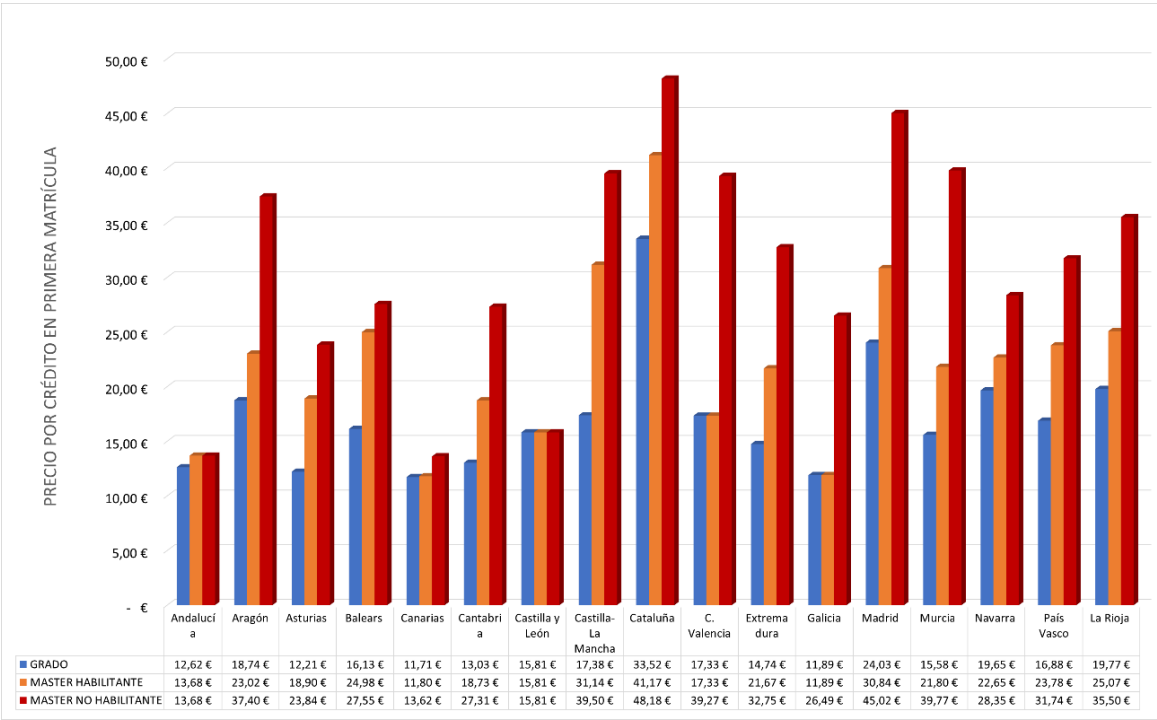


Figura 19: Precios públicos para españoles y comunitarios por CCAA y nivel

⁹ Máster necesario para ejercer en las 13 actividades profesionales reguladas específicas.

❖ Precios públicos para estudiantes extranjeros

En el caso de los extranjeros, para establecer los precios públicos pagados por crédito principalmente se toma como mínimo el precio de tercera matrícula por crédito correspondiente a un alumno español. En el caso de Máster habilitante se ha tomado en cuenta el precio de tercera matrícula, mientras que para máster no habilitante el de cuarta matrícula, lo que generalmente se hace en la mayoría de los centros. Al hablar de precios de doctorado, no se cobra por crédito sino por tutela doctoral y no existe diferencia de precios comparándolo con lo que pagan los estudiantes españoles o comunitarios.

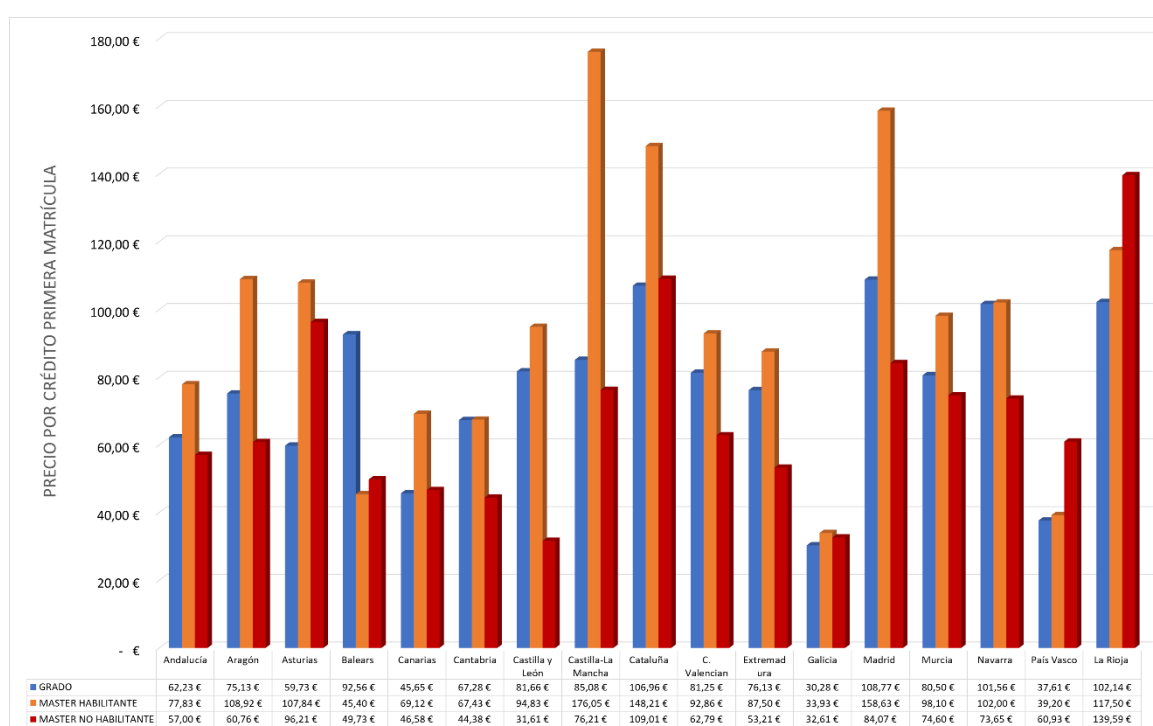


Figura 20: Precios públicos para estudiantes extranjeros por CCAA y nivel

Llama la atención la gran diferencia en la escala del eje Y de ambos gráficos, donde el límite superior en el caso de extranjeros es de 180€ por crédito, mientras que en el de estudiantes europeos el límite máximo es de 50€. Además, resulta interesante ver que en Baleares el precio por crédito de Grado es muy superior al de Máster sea habilitante o no. Por último y para establecer un ejemplo más claro, vemos en la figura 21, una comparación directa entre el precio promedio de cada nivel de estudio de un español o comunitario versus el de un estudiante extranjero o extracomunitario, recordando que el precio de doctorado es el mismo en ambos casos.

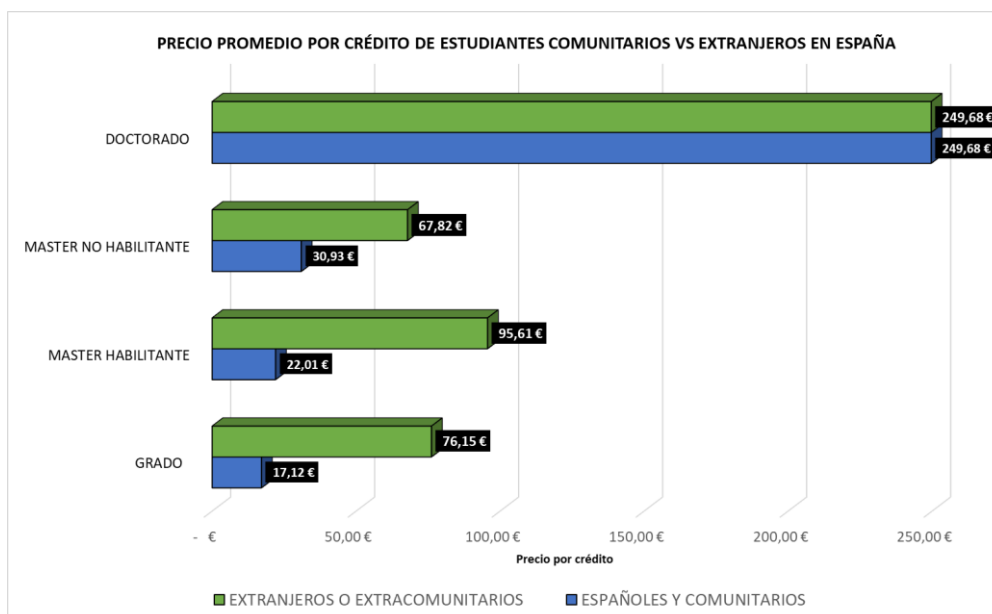


Figura 21: Comparativa de precios para estudiantes comunitarios vs extranjeros

La diferencia es muy alta comparando lo pagado por comunitarios frente a extracomunitarios, excluyendo el nivel de estudios de doctorado. En el caso de estudios de grado, el estudiante extranjero paga en promedio 76,15€, un 345% más por cada crédito en el que se matricula. En estudios de Máster habilitante paga en promedio 95,61€ por crédito, 334% más frente al estudiante comunitario y finalmente en estudios de Máster no habilitante paga 67,82€ por crédito, un 119% más que el estudiante español o comunitario. Es importante recordar que estos precios corresponden a Universidades públicas, cuyo objetivo es justamente ofrecer educación a precios asequibles para los ciudadanos españoles y europeos, por esto las diferencias son tan altas.

5.1.8.2 Precios de alquiler

Dentro del análisis de precios, resulta necesario realizar un balance de la variación del precio de alquiler para los estudiantes, muchos de ellos deciden el destino de estudios por temas económicos, donde no solo influye el precio de los estudios, sino el costo de vida, esto incluye transporte, alimentación, alquiler, etc. El principal factor que define el costo de vida viene a ser el precio de alquiler mensual, según una encuesta realizada por Uniplaces en 2018, el 75% de estudiantes extranjeros prefieren pagar un alquiler por una habitación en un piso compartido. Por lo que en este estudio se comparará los precios y su variación tomando en cuenta este hecho, así, se ha comparado el precio de alquiler de una habitación en piso compartido incluyendo las

Comunidades Autónomas con más oferta académica y por tanto las ciudades que albergan más estudiantes extranjeros. Los datos corresponden al año 2018 y han sido obtenidos de las estadísticas de Idealista, plataforma especializada en la venta, compra y alquiler de pisos.

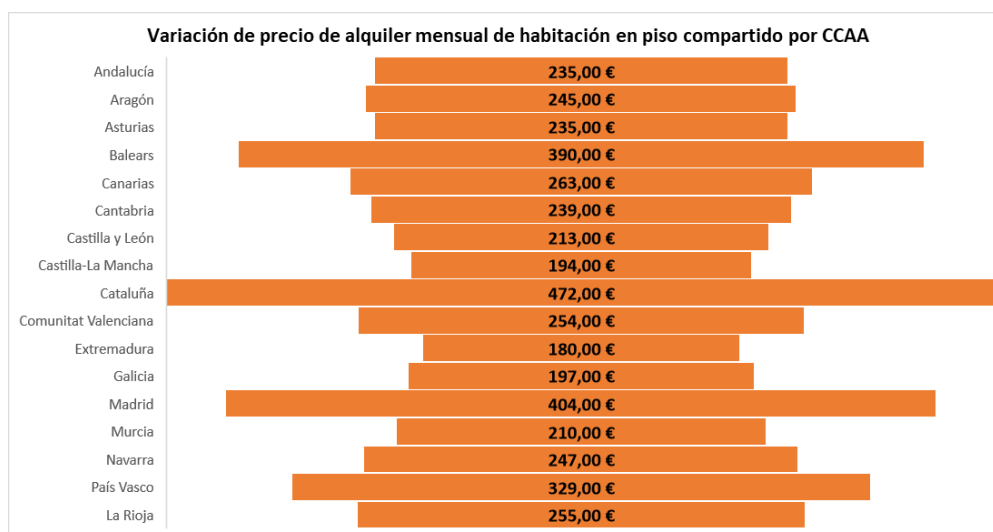


Figura 22: Precio de alquiler mensual por habitación en piso compartido por CCAA

Cataluña y Madrid vienen a ser las Comunidades con precios de alquiler por habitación más alto, superiores a los 400€ en ambos casos. Luego, llama la atención que Baleares ocupe el tercer lugar, en este caso, por un factor opuesto, al existir menos oferta. En cuarto lugar, con un precio promedio mensual que sigue siendo alto (329€) está el País Vasco, las demás Comunidades restantes mantienen precios similares que oscilan entre los 200€ y los 270€. Extremadura es la que presenta un menor precio promedio mensual por alquiler de una habitación siendo este de 180€. Para un análisis más específico, vemos a continuación las diferencias en 2018 de precio de alquiler de una habitación en las principales ciudades universitarias.

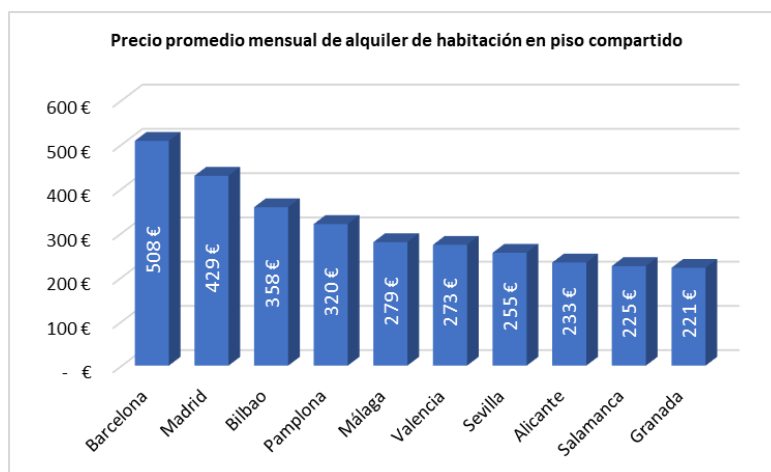


Figura 23: Precio de alquiler mensual de habitación en piso compartido por ciudad

5.2. Análisis de Correspondencias

En este estudio, se han realizado varios análisis de correspondencias que permitan encontrar relaciones entre variables y categorías importantes sobre los extranjeros que solicitan su estancia por estudios en España. Para ello, se ha utilizado el software SAS Base y aplicado principalmente el procedimiento *Proc Corresp* para obtener los resultados buscados.

5.2.1. Relación entre Edad y Región de procedencia

En primera instancia, analizaremos la relación entre la Edad de los estudiantes extranjeros, agrupada en 5 rangos distintos, y su región de procedencia. A continuación, vemos la tabla de contingencia que se usará para nuestro análisis.

Tabla de contingencia							
	RESTO_EUROPA	AFRICA	NORTEAMERICA	CENTROYSURAMERICA	ASIA	OCEANIA	Suma
18-24	1764	3466	4656	5723	9532	131	25272
25-29	653	1143	3115	9173	3341	79	17504
30-34	360	461	1159	5249	1281	30	8540
35-39	170	220	443	1822	536	11	3202
40-mas	120	163	374	1375	372	24	2428
Suma	3067	5453	9747	23342	15062	275	56946

Tabla 6: Tabla de contingencia entre variables edad y región de procedencia

Al observar la descomposición de la inercia y Chi-cuadrado, claramente debemos quedarnos con una sola dimensión, pues explica perfectamente la mayoría de la información, un 98,43% y el p-valor es bastante mayor a 0.05, siendo en este caso de 0,3542.

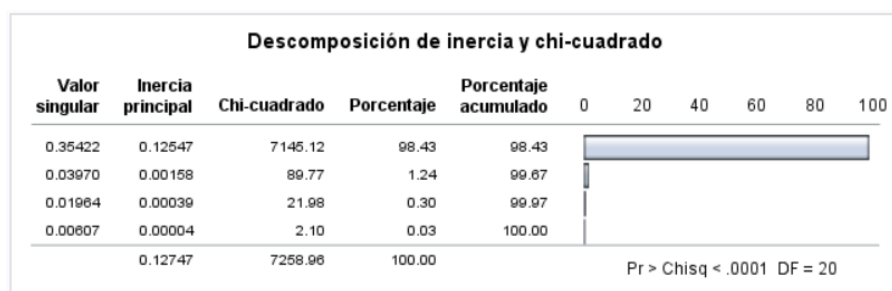


Figura 24: Descomposición del inercia y chi-cuadrado – Edad vs Región

Representando gráficamente las aportaciones y contribuciones al chi-cuadrado de las distintas relaciones tenemos una idea más clara de las que tienen mayor relevancia. En este

caso, en el mapa de calor a continuación vemos que mientras más oscuro es el tono de azul, mayor es la correlación entre el rango de edad y la región de procedencia.

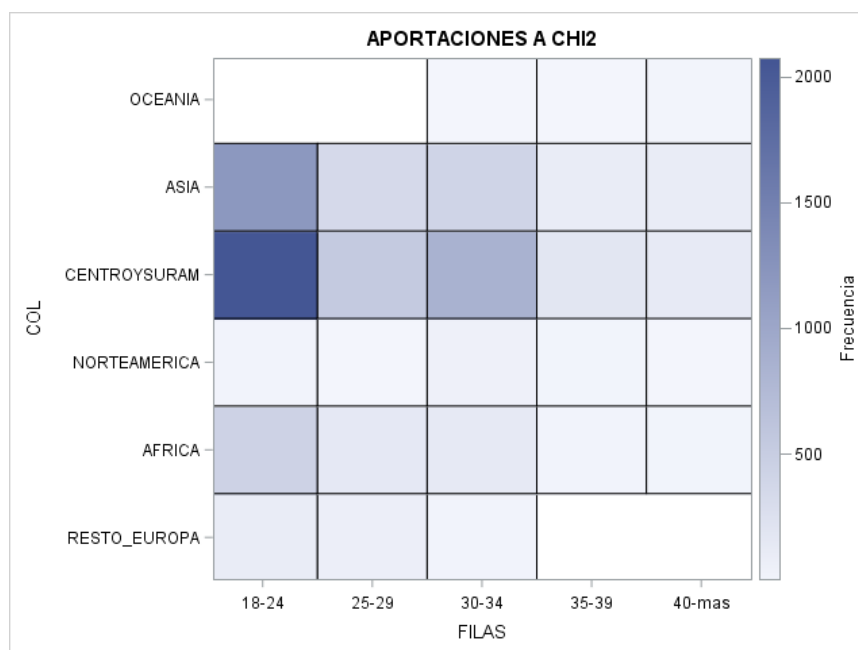


Figura 25: Mapa de calor de variables Edad y Región de procedencia

Los estudiantes procedentes de **Centro y Sur América** de entre **18-24 años** son los más correlacionados, luego están las personas de este mismo rango de edad provenientes de Asia y finalmente de África. A partir de estas relaciones más notorias, existe relación entre los estudiantes de 25 a 39 años de Asia y Centro y Sudamérica. En cuanto al resto de cruces de categorías, existen relaciones neutrales.

Si queremos ver de forma cuantitativa lo representado en el mapa de calor, a continuación, se muestra la tabla de contribuciones al chi-cuadrado total.

Contribuciones al estadístico chi-cuadrado total							
Porcentajes	RESTO_EUROPA	AFRICA	NORTEAMERICA	CENTROYSURAMERICA	ASIA	OCEANIA	Suma
18-24	1.643	6.229	0.348	28.581	16.712	0.009	53.522
25-29	1.227	2.336	0.065	7.666	4.942	0.005	16.241
30-34	0.299	2.144	0.864	12.031	5.831	0.042	21.212
35-39	0.000	0.337	0.277	2.725	1.572	0.018	4.930
40-mas	0.012	0.286	0.057	1.996	1.566	0.177	4.095
Suma	3.182	11.332	1.611	53.000	30.624	0.251	100.000

Tabla 7: Contribuciones al estadístico chi-cuadrado total Edad vs Región

- **Representación en el Plano factorial**

Tomando en cuenta las dos dimensiones que por defecto genera SAS y verificando que estas representan la mayor cantidad de información, generamos el plano factorial, que nos brinda una idea clara sobre las principales correlaciones entre las variables.

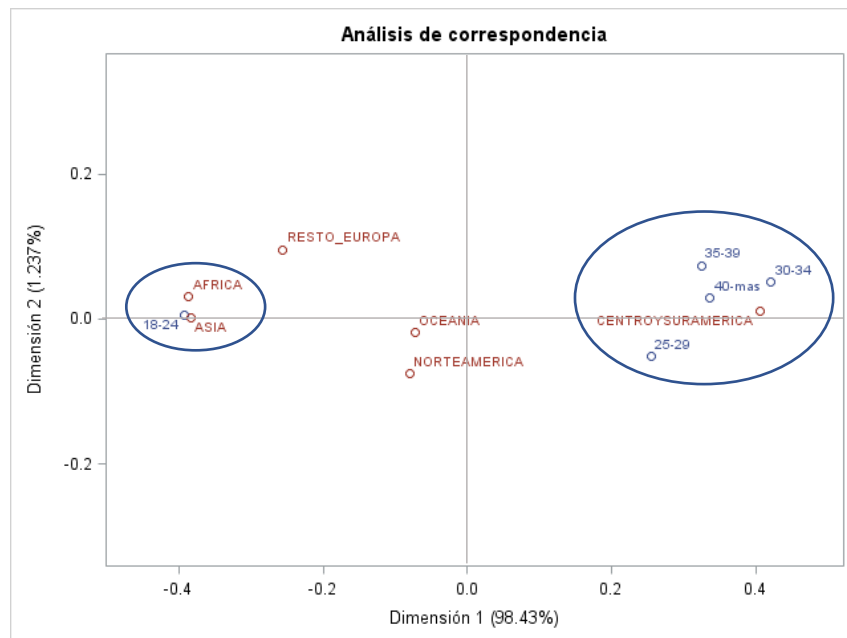


Figura 26: Plano factorial de variables Edad y Región de procedencia

Observamos que los estudiantes de Centro y Sur América tienen una correlación alta y destacan con la mayoría de los rangos de edad entre 25 a 29 años, 30 a 34 , 35 a 39 años y de 40 años o más, representan el 53% del total tomando en cuenta todas las edades. Además, la mayoría de los estudiantes extranjeros tienen entre 18 y 24 años, representando el 53% del total. Los estudiantes provenientes de Asia y África son en su mayoría de entre 18 y 24 años.

- **Análisis de perfiles**

Analizando por separados los perfiles de nuestra tabla de contingencia (Tabla 6). El perfil columna corresponde a las regiones de procedencia de los estudiantes, que son 6 distintas. Al representar gráficamente este perfil, observamos un gran despunte de la mayoría de las regiones entre los estudiantes de 18 a 24 años, donde en 5 de las 6 regiones proporcionalmente se tiene como máximo este rango. La única excepción es Centro y Sur América, siendo su punto máximo el de personas de una edad mayor, de entre 25 y 29 años.

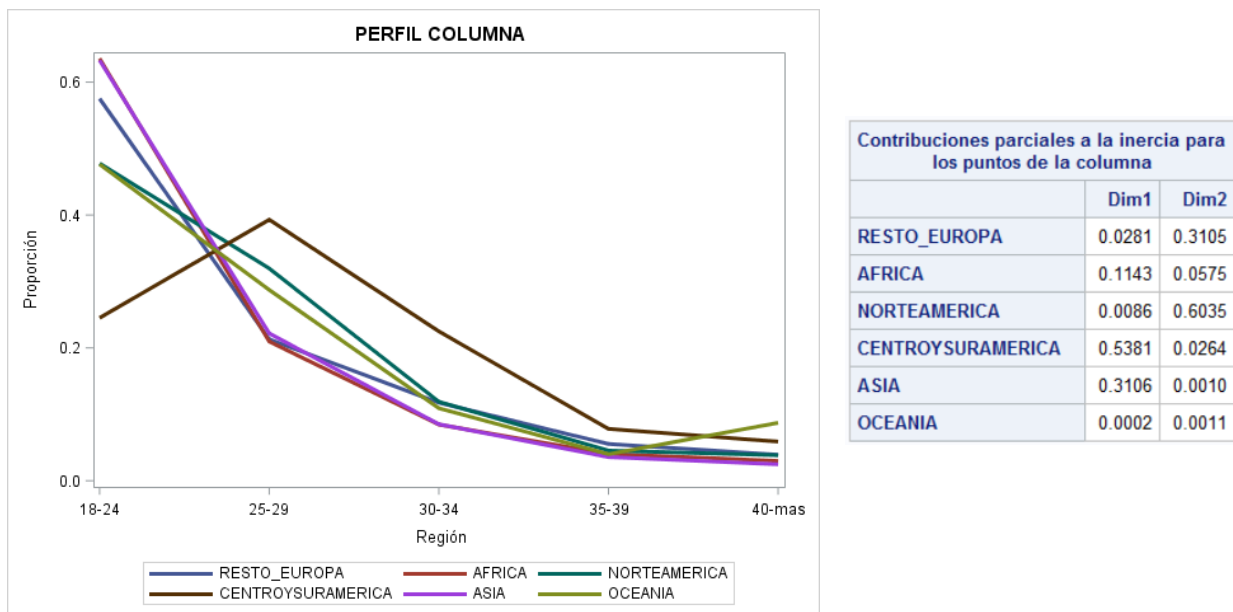


Figura 27: Representación y contribuciones del perfil columna

El perfil fila en cambio, se centra en los rangos de edad de los estudiantes extranjeros en España, claramente la mayoría de los grupos de edad, tienen su máximo en la región de Centro y Sur América. Sin embargo, tomando en cuenta los rangos de edad de entre 18 y 24 años, la mayor cantidad de estos son provenientes de Asia.

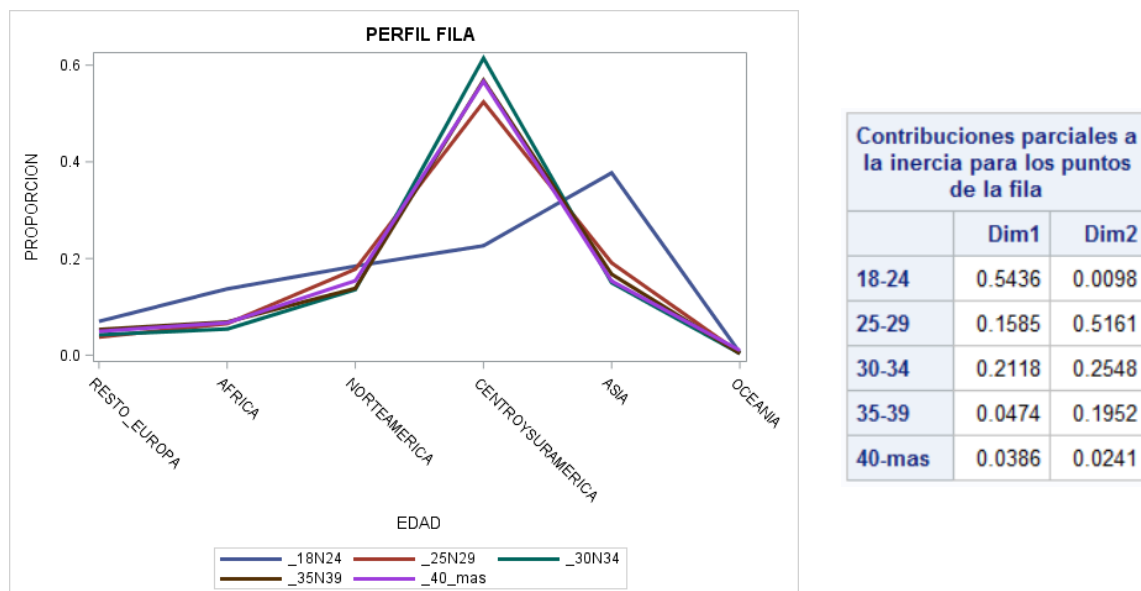


Figura 28: Representación y contribuciones del perfil fila

5.2.2. Relación entre Región de procedencia y CCAA de destino

El segundo análisis de correspondencias simple tiene como objetivo verificar la relación entre la región de Procedencia y la Comunidad Autónoma de destino de los estudiantes extranjeros. A continuación, se muestra la tabla de contingencia que se compone por 7 columnas y 16 filas. Se han agrupado algunas comunidades autónomas por características similares y por ubicación geográfica para tener datos más representativos e idóneos dentro del análisis. En este caso se ha unido a Asturias y Cantabria, y a Canarias y Baleares. Es importante mencionar que se toma en cuenta en este apartado a Ceuta y Melilla.

Procedimiento CORRESP							
Tabla de contingencia							
	RESTO_EUROPA	AFRICA	NORTEAMERICA	CENTROYSURAMERICA	ASIA	OCEANIA	Suma
ANDALUCÍ	514	1021	610	1825	1579	25	5574
ARAGÓN	32	161	64	248	319	4	828
ASTYCANT	25	43	64	410	68	5	615
CANARYBA	78	100	55	229	130	3	595
CAST_LEÓ	64	171	213	1012	1361	10	2831
CAST_LAM	22	63	50	196	229	1	561
CATALUÑA	1497	1017	1772	7867	4547	41	16741
C.VALENC	605	619	481	2441	1618	10	5774
EXTREMAD	8	41	132	125	17	6	329
GALICIA	44	74	105	361	375	1	960
MADRID	413	1129	2211	8324	5918	59	18054
MURCIA	53	94	63	307	212	1	730
NAVARRA	21	84	222	965	115	10	1417
PAISVASC	53	58	145	533	169	5	963
RIOJA	9	10	28	36	22	3	108
CEUTAYME	7	30	25	65	20	1	148
Suma	3445	4715	6240	24944	16699	185	56228

Tabla 8: Tabla de contingencia Región de procedencia y CCAA de destino

Al observar ahora la descomposición de inercia y chi-cuadrado, dos dimensiones no llegan a explicar la inercia en un porcentaje razonable, por lo que decidimos continuar el análisis con 3 dimensiones, explicando así el 95%.

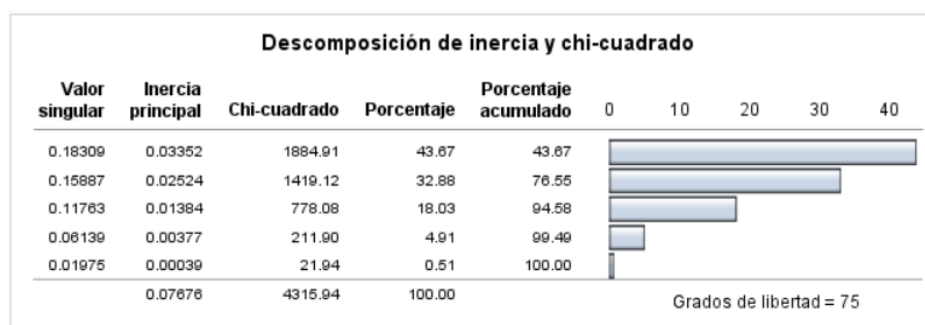


Figura 29: Descomposición de inercia y chi-cuadrado - Región vs CCAA.

El mapa de calor que se mostrará a continuación representa las aportaciones al Chi- y nos permite sacar conclusiones más rápidamente para ver entre qué categorías hay más relación.

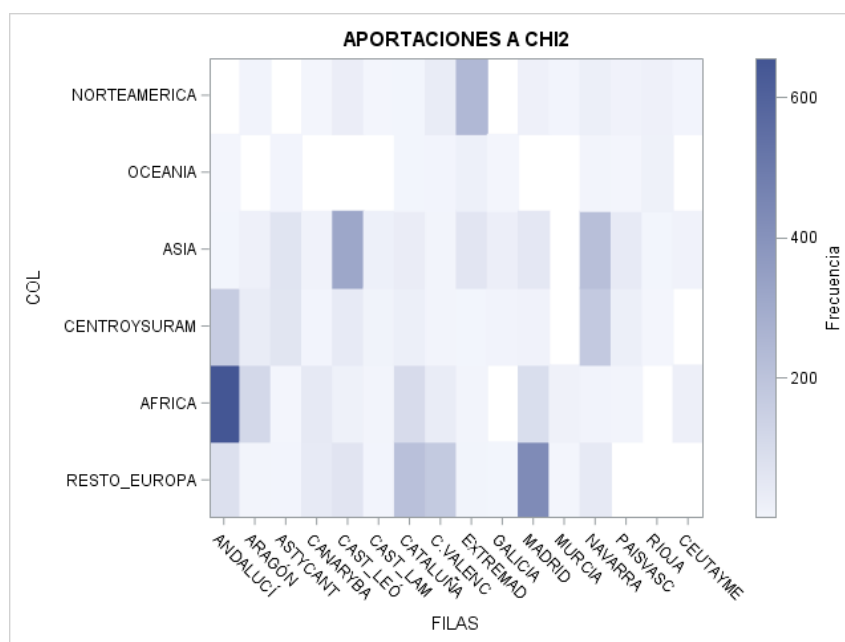


Figura 30: Representación del mapa de calor CCAA vs Región

Las categorías entre las que más relación existe son: *Andalucía-África* con una aportación de 15.19, en segundo lugar, tienen más relación las categorías *Madrid-Resto de Europa* con una aportación de 10,06 al chi cuadrado total, luego se relacionan más *Castilla y León-Asia*, con una aportación al chi cuadrado de 7,45 y en cuarto lugar está la relación entre *Extremadura-Norte América* que aporta 5,78 al chi cuadrado total. Ver tabla en Anexo B.

• Representación en el Plano factorial

A continuación, se observa el plano factorial que compara la dimensión 1 y 2, se toma este como el principal plano factorial donde se observan mejor las relaciones entre las distintas categorías de las variables. A través de su interpretación, podemos sacar las siguientes conclusiones:

- Los estudiantes de Centro y Sur América no tienen una preferencia de CCAA de destino, se distribuyen por todas ellas, siendo las más cercanas Madrid, Cataluña y Murcia.
- Los estudiantes provenientes de Asia también se distribuyen por todo el país, teniendo correlación más estrecha con Galicia, luego Madrid, Castilla León y Castilla la Mancha.
- Los estudiantes de África están representados en la dimensión 1 y relacionados con Andalucía, y luego la Comunidad Valenciana.

- Los estudiantes de Norteamérica tienden a optar por estudiar en Cataluña.
 - Los provenientes de Oceanía principalmente se distribuyen entre Cataluña y la Rioja.
- Finalmente, los del resto de Europa destacan más en la dimensión 1 por elegir Comunidades autónomas como Canarias y Baleares.

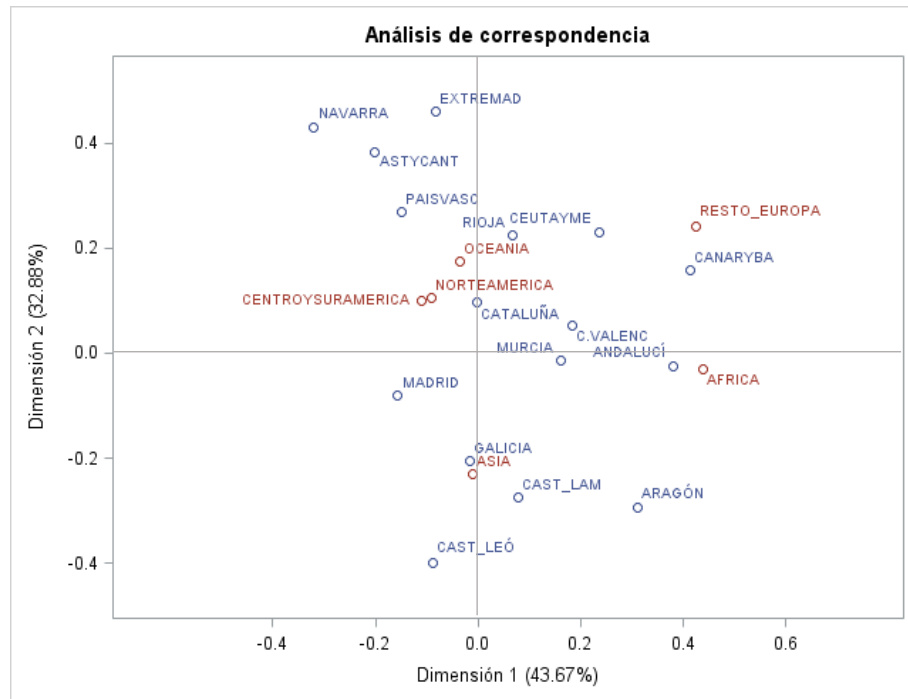


Figura 31: Plano factorial de variables CCAA y Región de procedencia

5.3. Modelos de Regresión

En nuestro caso, debido a que nuestras variables tienen problemas de multicolinealidad al estar bastante relacionadas, procederemos a aplicar dos técnicas de regresión avanzada para corregir esto y encontrar la verdadera influencia y la importancia de las variables en relación con la cantidad de estudiantes extranjeros en España.

5.3.1. Regresión LASSO

En primer lugar, aplicaremos la Regresión LASSO para ver la influencia de las siguientes variables sobre la variable dependiente Y, que en este caso es Estudiantes, refiriéndonos a la cantidad de estudiantes extranjeros en España y tomando en cuenta como individuos cada una de las Comunidades Autónomas.

CCAA	ESTUDIANTES EXTRANJEROS (Y)	UNIVERSIDADES PÚBLICAS	UNIVERSIDADES PRIVADAS	PRECIO POR CRÉDITO				PRECIO POR HABITACIÓN EN PISO COMPARTIDO
				GRADO	MASTER HABILITANTE	MASTER NO HABILITANTE	TUTELA DOCTORADO	
Andalucía	6441	9	1	62,23 €	77,83 €	57,00 €	60,30 €	235,00 €
Aragón	867	1	1	75,13 €	108,92 €	60,76 €	218,95 €	245,00 €
Asturias	406	1	0	59,73 €	107,84 €	96,21 €	200,00 €	235,00 €
Baleares	326	1	0	92,56 €	45,40 €	49,73 €	210,00 €	390,00 €
Canarias	567	2	2	45,65 €	69,12 €	46,58 €	203,80 €	263,00 €
Cantabria	491	1	1	67,28 €	67,43 €	44,38 €	206,60 €	239,00 €
Castilla y León	3844	4	4	81,66 €	94,83 €	31,61 €	225,89 €	213,00 €
Castilla-La Mancha	847	1	0	85,08 €	176,05 €	76,21 €	400,85 €	194,00 €
Cataluña	16911	7	4	106,96 €	148,21 €	109,01 €	401,12 €	472,00 €
C. Valencia	6373	5	3	81,25 €	92,86 €	62,79 €	300,00 €	254,00 €
Extremadura	326	1	0	76,13 €	87,50 €	53,21 €	104,00 €	180,00 €
Galicia	1281	3	0	30,28 €	33,93 €	32,61 €	200,00 €	197,00 €
Madrid	16758	6	7	108,77 €	158,63 €	84,07 €	390,00 €	404,00 €
Murcia	877	3	1	80,50 €	98,10 €	74,60 €	388,00 €	210,00 €
Navarra	1434	2	1	101,56 €	102,00 €	73,65 €	223,35 €	247,00 €
País Vasco	1188	3	2	37,61 €	39,20 €	60,93 €	204,50 €	329,00 €
La Rioja	208	1	0	102,14 €	117,50 €	139,59 €	207,26 €	255,00 €

Tabla 9: Variables tomadas en cuenta para la regresión LASSO

En primer lugar, verificamos la correlación lineal entre las variables, vemos en la siguiente tabla que las variables con más correlación con la variable dependiente son el número de Universidad públicas y Universidades Privadas, y el Precio de Alquiler de una habitación.

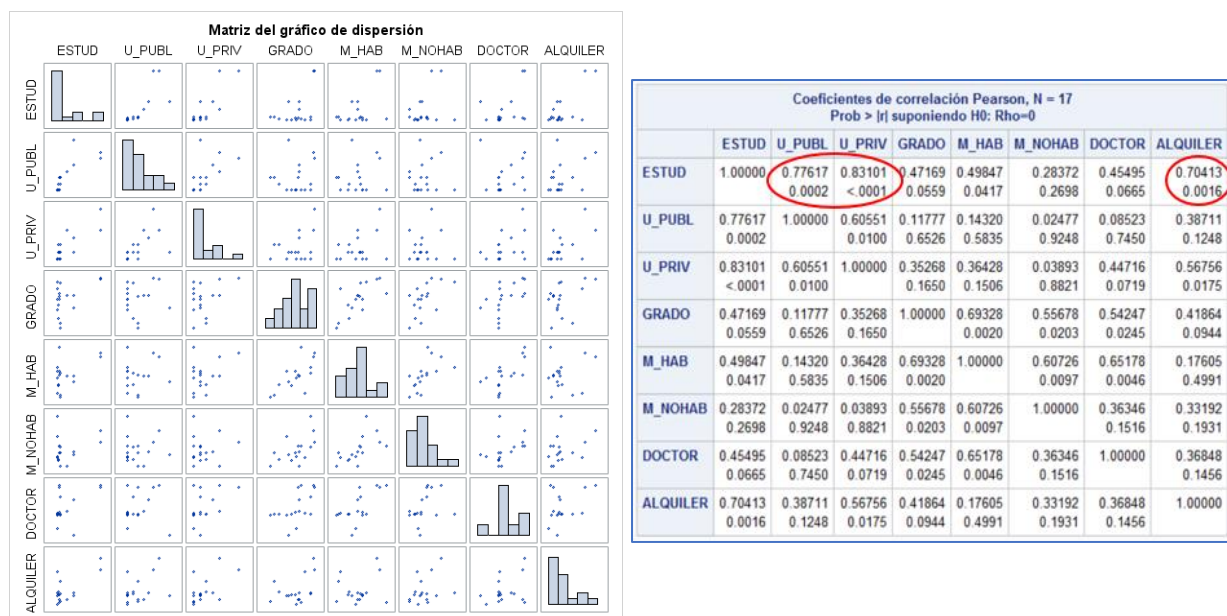


Figura 32: Correlación lineal de las variables

Después de verificar la correlación y ver las variables que tienen más influencia sobre el número de estudiantes extranjeros, aplicamos la Regresión Lasso con criterio SBC, aunque los demás criterios como AIC, R^2 , etc. han elegido el mismo modelo, obteniendo los siguientes resultados y eligiendo el modelo en el paso 4, incluyendo estas variables.

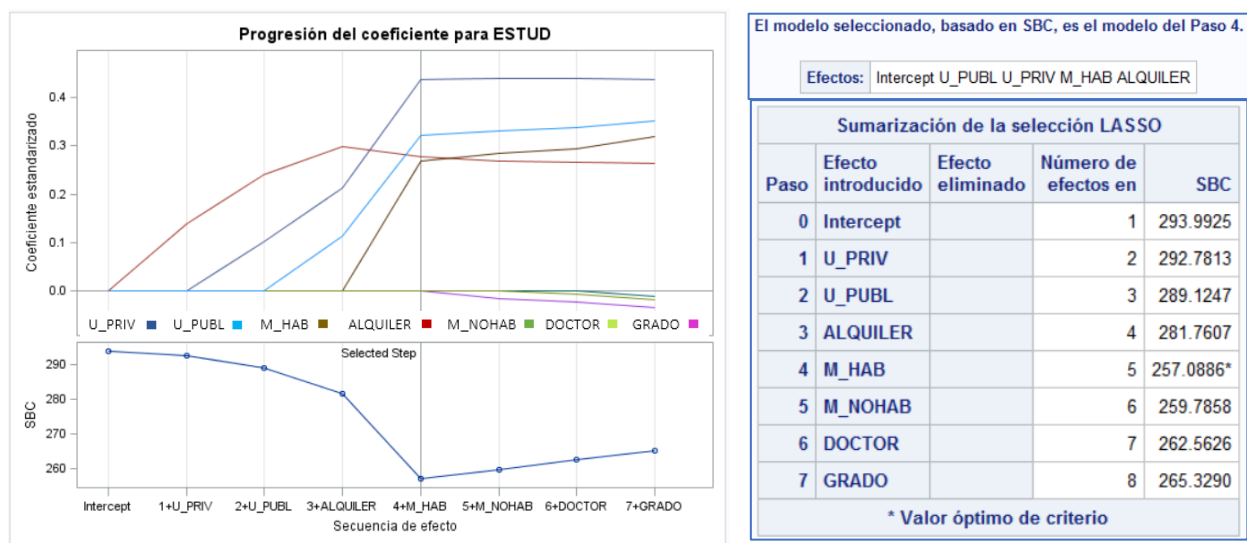


Figura 33: Modelo seleccionado aplicando Regresión LASSO

Las variables elegidas por el modelo son las mismas que vimos inicialmente que tienen una correlación más alta. Sin duda alguna, la cantidad de Universidades Públicas y Privadas es lo que más influye para que haya una mayor cantidad de estudiantes extranjeros en una Comunidad Autónoma, aunque el resultado parece obvio es importante corroborarlo y por ello se ha establecido este análisis.

En cuanto al alquiler y su influencia sobre la variable dependiente, decimos que este es un caso clásico de regresión espuria, en donde esta influencia que se refleja en realidad es debida a las dos anteriores variables, pues el número de universidades que hay en las ciudades grandes hacen que los precios del alquiler sean mayores. Finalmente, al referirnos a la cuarta variable del precio por crédito del Máster Habilitante, esto si llama la atención, pues aun siendo precios tasados, existen horquillas o rangos, que en las CCAA como Madrid y Barcelona permiten tener máximos más altos y por ello su influencia en el número de estudiantes extranjeros, a más caro estos masters y a más caro el alquiler, mayor número de estudiantes extranjeros habrá.

Observamos ahora los resultados del procedimiento aplicado en SAS Base y los principales indicadores estadísticos y medidas de adecuación del modelo elegido por Regresión LASSO y aplicando el criterio de selección de SBC. Vemos que los demás criterios como el BIC, AIC, son bastante cercanos al seleccionado y de igual manera el R^2 y el R^2 ajustado tienen valores altos y bastante cercanos a 1.

Raíz MSE	1508.59580
Media dependiente	3479.11765
R-cuadrado	0.9414
R-Sq Ajust	0.9219
AIC	271.92256
AICC	280.32256
BIC	262.04534
C(p)	2.13830
SBC	257.08863
ASE	1606490

Estimadores de parámetros		
Parámetro	DF	Estimador
Intercept	1	-9804.569147
U_PUBL	1	962.157160
U_PRIV	1	772.063312
M_HAB	1	36.513393
ALQUILER	1	21.166135

Análisis de la varianza				
Fuente	DF	Suma de cuadrados	Cuadrado de la media	F-Valor
Modelo	4	438937792	109734448	48.22
Error	12	27310336	2275861	
Total corregido	16	466248128		

Tabla 10: Resultados del modelo seleccionado por Regresión LASSO

Es importante mencionar que, a partir de estos resultados, podríamos obtener predicciones de la variable estudiantes extranjeros, tomando en cuenta los resultados de la Regresión LASSO, sin embargo, al tener pocas observaciones no vale la pena hacerlo, pues las predicciones son muy poco acertadas y nuestro objetivo realmente era evaluar la influencia que tienen los predictores sobre la variable dependiente.

5.3.2. Regresión PLS

La segunda regresión aplicada en este trabajo será la PLS o de Mínimos cuadrados parciales, que nos permitirá hacer frente a la multicolinealidad de nuestros datos haciendo una regresión multivariable, en este caso tendremos tres variables dependientes $Y(Y_1, Y_2, Y_3)$, por lo que aplicaremos una regresión PLS2. Las variables utilizadas para este procedimiento son algunas de las que se usaron en la regresión LASSO, sin embargo, aquí las variables dependientes son tres y se aumenta la región de procedencia de los estudiantes, lo que será interesante para saber el nivel de influencia de las variables predictoras en cada una de las dependientes. Es importante mencionar que aquí no se analizan los factores, sino que se obtienen para analizar e intentar aumentar la correlación entre las variables X e Y .

CCAA	DEPENDIENTES Y1,Y2,Y3			PRECIO POR CRÉDITO INSCRITO PRIMERA MATRÍCULA (€)			
	ESTUDIANTES DE GRADO	ESTUDIANTES DE MÁSTER	ESTUDIANTES DE DOCTORADO	GRADO	MASTER HABILITANTE	MASTER NO HABILITANTE	TUTELA DOCTORADO
Andalucía	3.537	2.329	2.565	62,23 €	77,83 €	57,00 €	60,30 €
Aragón	510	140	283	75,13 €	108,92 €	60,76 €	218,95 €
Asturias	123	179	188	59,73 €	107,84 €	96,21 €	200,00 €
Baleares	286	94	153	92,56 €	45,40 €	49,73 €	210,00 €
Canarias	697	98	173	45,65 €	69,12 €	46,58 €	203,80 €
Cantabria	457	564	50	67,28 €	67,43 €	44,38 €	206,60 €
Castilla y León	1.862	2.673	1.196	81,66 €	94,83 €	31,61 €	225,89 €
Castilla-La Mancha	291	179	227	85,08 €	176,05 €	76,21 €	400,85 €
Cataluña	6.919	8.768	4.489	106,96 €	148,21 €	109,01 €	401,12 €
C. Valencia	3.745	2.492	1.970	81,25 €	92,86 €	62,79 €	300,00 €
Extremadura	99	76	126	76,13 €	87,50 €	53,21 €	104,00 €
Galicia	515	359	702	30,28 €	33,93 €	32,61 €	200,00 €
Madrid	7.506	8.495	3.163	108,77 €	158,63 €	84,07 €	390,00 €
Murcia	1.147	355	300	80,50 €	98,10 €	74,60 €	388,00 €
Navarra	1.074	984	331	101,56 €	102,00 €	73,65 €	323,35 €
País Vasco	762	601	571	37,61 €	39,20 €	60,93 €	204,50 €
La Rioja	103	15	22	102,14 €	117,50 €	139,59 €	207,26 €

PRECIO POR HABITACIÓN EN PISO COMPARTIDO	REGIÓN DE PROCEDENCIA (total)					
	Resto de Europa	África	América del Norte	América Central y del Sur	Asia	Oceanía
235,00 €	514	1021	610	1825	1579	25
245,00 €	32	161	64	248	319	4
235,00 €	9	23	27	108	48	2
390,00 €	34	19	27	74	51	1
263,00 €	44	81	28	155	79	2
239,00 €	16	20	37	302	20	3
213,00 €	64	171	213	1012	1361	10
194,00 €	22	63	50	196	229	1
472,00 €	1497	1017	1772	7867	4547	39
254,00 €	605	619	481	2441	1618	10
180,00 €	8	41	132	125	17	6
197,00 €	44	74	105	361	375	1
404,00 €	413	1129	2211	8324	5918	59
210,00 €	53	94	63	307	212	1
247,00 €	21	84	222	965	115	10
329,00 €	53	58	145	533	169	4
255,00 €	9	10	28	36	22	3

Tabla 11: Variables tomadas en cuenta para la Regresión PLS

El primer paso es seleccionar el número de factores basándonos en el PRESS, el procedimiento inicial sugiere 4 factores, sin embargo, realizando el procedimiento en SAS vemos que dos factores son suficientes, explicando los datos en un porcentaje mayor al 78%.

Cruce Validación para el número de factores extraídos			
Número de factores extraídos	PRESS de la media de la raíz	T**2	Prob > T**2
0	1.0625	7.899908	0.0030
1	0.412969	5.339174	0.1330
2	0.324164	2.968692	0.4430
3	0.330017	3.181777	0.4070
4	0.320102	0	1.0000
5	0.364513	5.322055	0.0620
6	0.382632	3.346762	0.3170
7	0.401682	7.620549	0.0170
8	0.452639	4.493745	0.1440
9	0.503765	3.656571	0.2850
10	0.53238	4.901176	0.0900
11	0.689723	6.151494	0.0150

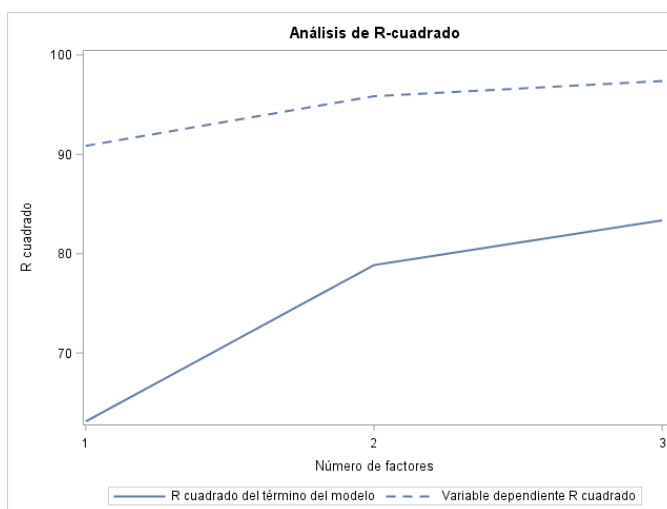


Figura 34: Número de factores extraídos aplicando PLS

Observamos en la siguiente tabla, el porcentaje explicado por los factores, con el primer y segundo factor, logramos explicar las variables dependientes en más del 95%, están muy bien explicadas. Mientras que las variables predictoras o las X en nuestro modelo están explicadas en un 78.85%, valor que podría mejorar pero que es suficiente en este caso.

Variación del porcentaje explicado por Factores de cuadrados mínimos parciales													
Número de factores extraídos	Efectos del modelo												
	GRADO	M_HAB	M_NOHAB	DOCTOR	ALQUILER	RESTEUROPA	AFRICA	AM_NORTE	AM_CENTRALYSUR	ASIA	OCEANIA	Actual	Total
1	35.2254	35.6144	17.0259	28.4616	55.2275	67.9819	79.1209	95.5379	97.1907	93.7820	89.0675	63.1123	63.1123
2	74.3856	74.8152	61.7159	57.6645	55.3199	72.2616	89.6071	96.8037	97.9865	95.5431	91.2923	15.7418	78.8541

Variables dependientes				
ESTGRADO	ESTMASTER	ESTDOCT	Actual	Total
93.3403	94.9777	84.2670	90.8617	90.8617
97.4074	96.4311	93.7339	4.9958	95.8575

Tabla 12: Variación explicada por los factores – PLS

Ahora, vemos la representación de los perfiles de los estimadores estandarizados. Es de suma importancia que estén estandarizados para que todos los datos tengan media 0 y desviación típica 1, pues así obtenemos el efecto real de cada variable independiente de las unidades de medida, tomando en cuenta que algunas variables representan precios mientras que otras la cantidad de alumnos.

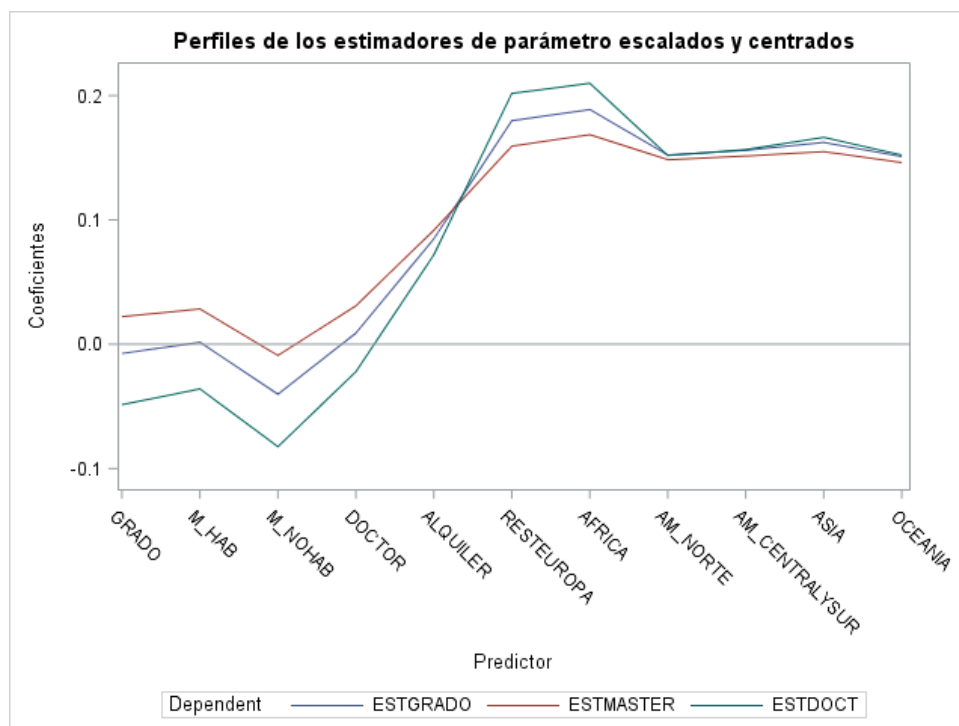


Figura 35: Estimadores de parámetros del modelo de regresión PLS

Los precios de los créditos de grados y del máster no habilitante no influyen en el número de estudiantes de grado y máster. Sin embargo, sí influyen de manera negativa en el número de estudiantes de doctorado. Por otra parte, es importante resaltar que existe el mayor número de estudiantes de África y Resto de Europa en estudios de doctorado.

En cuanto a la influencia del precio del alquiler que quizás se vea mejor en la regresión Lasso realizada anteriormente, es interesante fijarnos que a mayor precio de alquiler más estudiantes extranjeros en todos los estudios, este es un caso clásico de **regresión espuria**, en donde esta influencia que se refleja en realidad es debida a una tercera variable que es el número de número de universidades que hay en las ciudades grandes donde a su vez los precios del alquiler son mayores. A continuación, vemos cuantitativamente los estimadores de parámetros del modelo estandarizados.

Estimadores de parámetro para datos centrados y escalados			
	ESTGRADO	ESTMASTER	ESTDOCT
Intercept	0.0000000000	0.0000000000	0.0000000000
GRADO	-.0074278088	0.0222019421	-.0486429686
M_HAB	0.0016316890	0.0283671046	-.0358709144
M_NOHAB	-.0402277552	-.0090272650	-.0824088612
DOCTOR	0.0089571505	0.0309205614	-.0221381155
ALQUILER	0.0848114522	0.0917162762	0.0719515007
RESTEUROPA	0.1797716274	0.1592616091	0.2017324610
AFRICA	0.1887556650	0.1685810421	0.2099088637
AM_NORTE	0.1522410132	0.1484621584	0.1518064967
AM_CENTRALYSUR	0.1559699785	0.1513733089	0.1565404936
ASIA	0.1622963362	0.1549531167	0.1664752509
OCEANIA	0.1509865791	0.1460769899	0.1521827688

Tabla 13: Estimadores de parámetros de regresión PLS

Finalmente, es importante mencionar que con este procedimiento logramos solucionar la multicolinealidad, calculando unos factores de manera que exista más correlación entre las variables X y las variables Y. Repitiendo este proceso, hallando otros factores y regresiones lineales hasta lograr explicar de mejor manera los datos, maximizando así la correlación entre las variables Y y las X. Es importante tomar en cuenta que estos factores no son interpretables, sino que son justamente una herramienta para maximizar la correlación.

5.4. Modelos de Predicción

5.4.1. Análisis de la Serie temporal

La serie que se va a analizar a continuación muestra el número de estudiantes extranjeros que han solicitado autorización de estancia por estudios en España durante los últimos 26 años, desde el 1 de enero de 1993 hasta el tercer trimestre del 2019. Se trata de una serie temporal de baja frecuencia, pues presenta datos trimestrales. El análisis de esta resultará de mucha importancia para este trabajo porque nos permitirá hacer una predicción a futuro y además ver cómo ha ido evolucionando la cantidad de estudiantes extranjeros que eligen a España como su principal destino. Para ver el código utilizado en SAS ver Anexo A.

A continuación, se muestra la representación de la serie temporal observando las variaciones y su comportamiento en cuanto a tendencia, estacionalidad y los periodos.

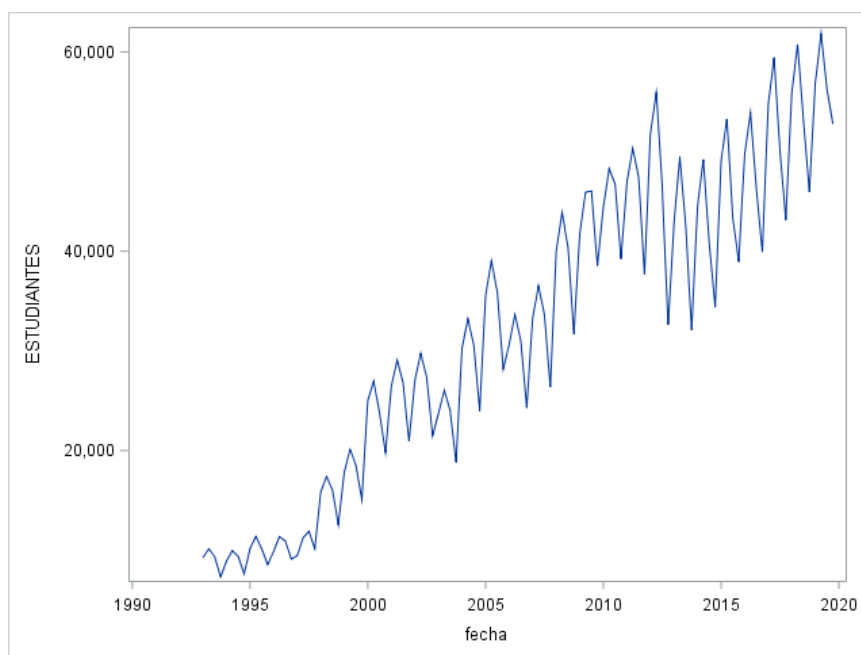


Figura 36: Representación de la serie temporal entre 1993 y 2019

Observamos que la serie tiene una tendencia de crecimiento constante desde 1993, alcanzando su punto más alto en el 2019. Para una descripción más específica, vemos su comportamiento estacional al representarla en un periodo menor en la siguiente figura.

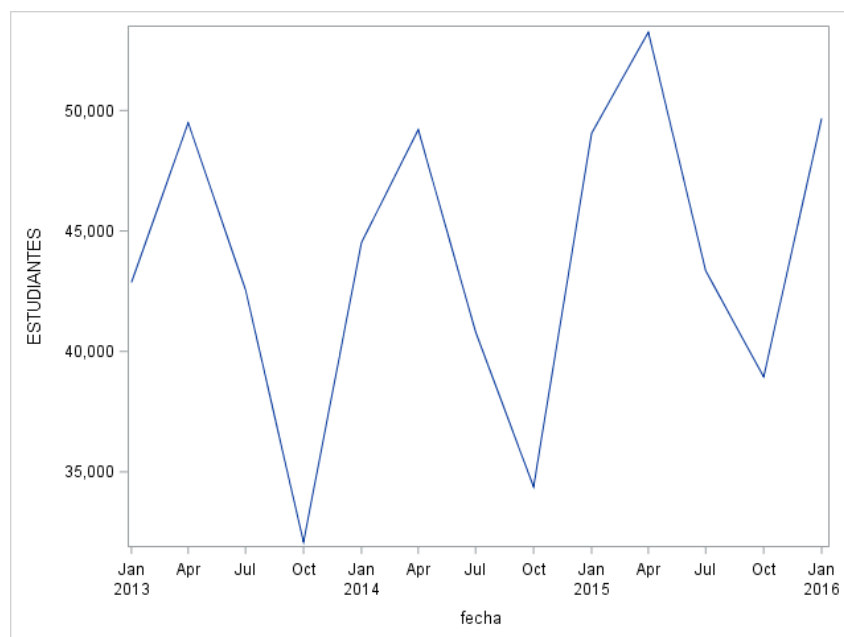


Figura 37: Comportamiento trimestral de la serie entre 2013 y 2016

El número de estudiantes extranjeros en España aumenta considerablemente entre el mes de octubre y abril (pico máximo), luego tiene una disminución brusca entre los meses de mayo a septiembre, con lo que intuimos que la serie tiene un componente estacional, procederemos entonces con la descomposición estacional de la misma.

Descomposición estacional

Después de realizar la descomposición estacional, si observamos los principales estadísticos obtenidos mediante el procedimiento *TIMESERIES*, decimos que el trimestre donde más estudiantes extranjeros obtienen su estancia por estudios en España es el segundo, que comprende los meses entre abril y julio con una media de 36.227 estudiantes. Mientras que el trimestre en donde se obtienen menos permisos de estancia por estudios en España es el cuarto de cada año, comprendido entre los meses de octubre y enero.

Estadísticos estacionales para la variable ESTUDIANTES						
Índice estacional	N	Mínimo	Máximo	Suma	Media	Desviación estándar
1	27	8917.000	56951.00	892159.0	33042.93	16013.157
2	27	9969.554	61828.28	978124.6	36226.84	17291.808
3	27	9352.823	56228.00	878870.8	32550.77	14929.534
4	27	7334.026	52768.00	721368.6	26717.36	12979.184

Tabla 14: Principales estadísticos estacionales de la serie

En la figura 38, vemos el gráfico generado por la descomposición estacional, la gráfica superior izquierda nos muestra la tendencia (Trend-Cycle) que en este caso no llega a ser cero por lo que existe una clara falta de estacionariedad en la serie. Mientras que, en el segundo gráfico, parte superior derecha (Seasonal_Irregular), vemos la representación de la estacionalidad o la serie sin tendencia. Luego vemos la componente irregular (Irregular) de nuestra serie y la serie ya desestacionalizada (Seasonally Adjusted). Se ve claramente que la serie tiene una tendencia positiva constante (ver tabla en el Anexo B).

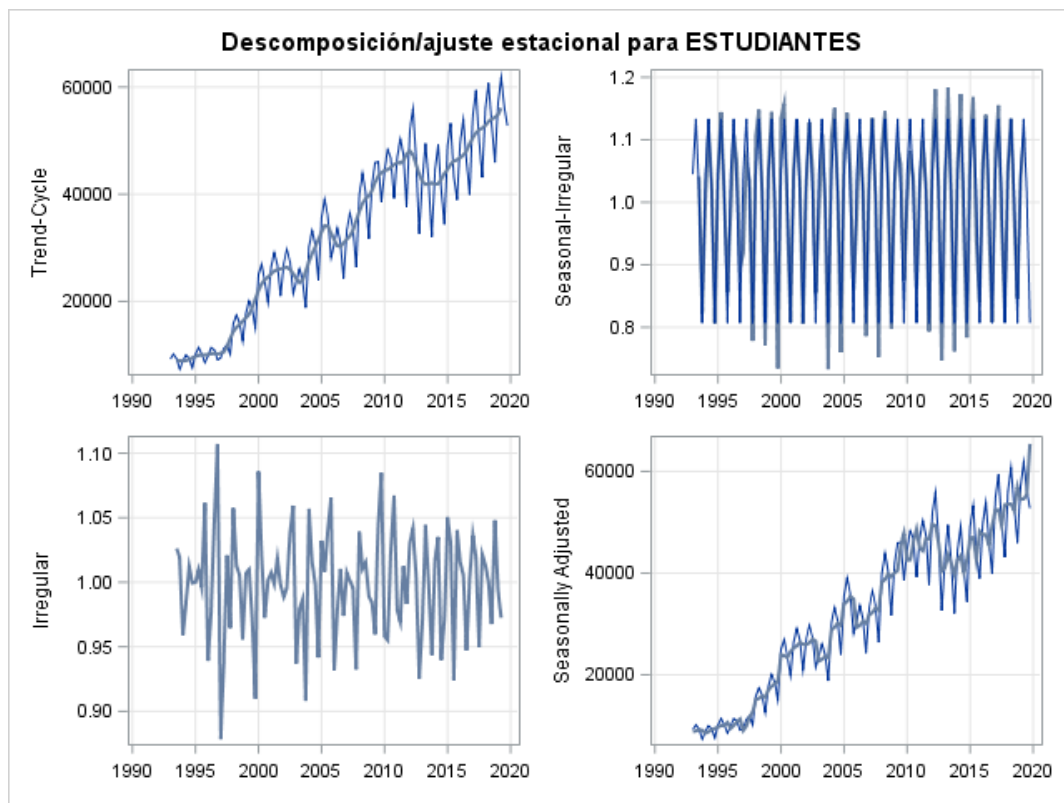


Figura 38: Representación de la descomposición estacional de la serie

Una vez analizada la serie con más detalle y observando su comportamiento en cuanto a tendencia y estacionalidad, emplearemos algunos de los métodos descritos en la teoría para ajustar la serie y realizar predicciones futuras. Para ello, se han reservado los datos reales de la serie del último año, para que estos sean el punto de comparación de las predicciones de cada método aplicado.

5.4.2. Aplicación de modelos de predicción

Aplicación de modelo de suavizado exponencial

El primer modelo por aplicar será un suavizado exponencial usando el Modelo multiplicativo de Holt-Winters, al presentar nuestra serie estacionalidad, este método hará un triple alisado con el que podremos hacer predicciones eliminando las fluctuaciones aleatorias y manteniendo solo el componente tendencia-estacionalidad que es el que interesa y del que se muestran los resultados en la siguiente tabla.

Método de Winters (Multiplicativo) Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.85362	0.06499	13.13	<.0001
Peso de tendencia	0.0010000	0.01272	0.08	0.9375
Peso estacional	0.58190	0.31192	1.87	0.0650

Tabla 15: Suavizado de Holt-Winters, método multiplicativo

Al fijarnos en los estimadores, el peso de nivel o constante es el que más variación tiene, de 0,85, lo que nos dice que por cada cambio en el tiempo (t-1) la constante varía en 0,85. En segundo lugar, el peso estacional es de 0,58, la estacionalidad sí varía en esa cantidad respecto al momento anterior. Lo que si se mantiene cercano a 0 es la pendiente o tendencia que no cambia en el tiempo pues su peso es 0,001.

De igual manera, al fijarnos en el p-valor, el parámetro de estacionalidad es significativamente distinto de cero por lo que **se rechaza la hipótesis nula** y decimos que la **serie es efectivamente estacional**. Como podemos ver en los resultados de su predicción, el modelo es bastante acertado.

Predicciones para la variable ESTUDIANTES					
Obs	Tiempo	Predicciones	Error estándar	Límites de confianza al 95%	
105	2019:1	59512.7818	2200.6732	55199.5416	63826.0220
106	2019:2	64527.3122	2988.9596	58669.0590	70385.5654
107	2019:3	57064.2148	3274.1369	50647.0245	63481.4052
108	2019:4	47289.5751	3348.7735	40726.0996	53853.0506

Tabla 16: Predicciones de estudiantes extranjeros en España, método Holt-Winters

Se observa en la tabla 16 que el error estándar de cada trimestre es bajo, lo que se refleja exactamente en los límites de confianza de la predicción, ya que sus rangos son aceptables, siendo en el primer trimestre de aproximadamente 8.600 estudiantes, mientras que en el último trimestre de 13100 estudiantes. El método multiplicativo de Holt-Winters permite predecir nuestra serie de buena manera, sus resultados se muestran gráficamente en la siguiente figura.

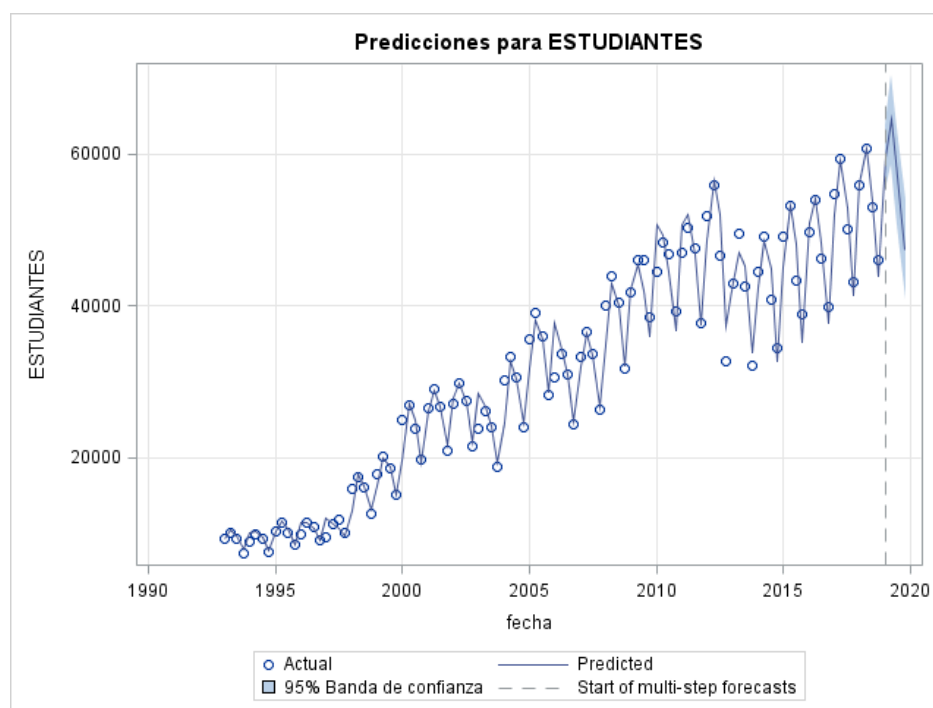


Figura 39: Serie original suavizada y predicciones de Holt-Winters

Se observa claramente como la serie suavizada se acopla muy bien a la serie original, dejan pocos puntos fuera de las líneas, lo que nos dice que la predicción y el método aplicado ha sido efectivo. Además, a partir de la línea discontinua, se observa la predicción obtenida mediante el suavizado de Holt-Winters y que esta tiene características muy similares al compararla con los mismos periodos de años pasados. Vemos a continuación los estadísticos de ajuste del suavizado.

Estadísticas de ajuste para la variable ESTUDIANTES	
Estadístico	Valor
Grados de error de libertad	101
Número de observaciones	104
Criterio de información de Akaike	1603.83175
Criterio de información bayesiano de Schwarz	1611.76492
R-cuadrado	0.97899714
R-cuadrado ajustado	0.97858124

Tabla 17 : Estadísticos de ajuste del suavizado Holt-Winters

Aplicación de modelo ARIMA con metodología Box Jenkins

Aplicaremos ahora el modelo ARIMA, como primer paso, representamos en la siguiente figura la serie para ver sus autocorrelogramas simple y parcial, lo que hacemos a través de *PROC ARIMA* de SAS. Podemos observar que la autocorrelación simple parcial (ACF) de la serie decrece, pero no de manera exponencial como se espera. A la par se observa en el gráfico de la parte superior izquierda, que la serie no es estacionaria, pues su media no es estable. Finalmente, sus autocorrelogramas parciales nos muestran que hay coeficientes fuera de las bandas, lo que nos lleva a concluir que se requerirá de una diferenciación de la serie, para que sea estacionaria.

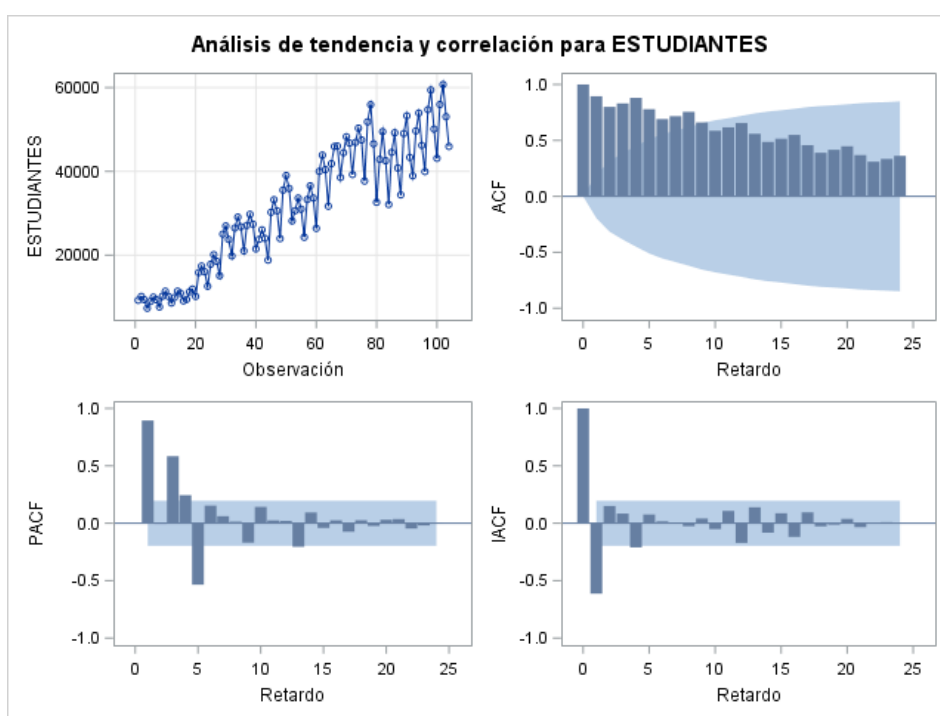


Figura 38: Representación de la función de autocorrelación Simple (ACF) y Parcial de la Serie (PACF)

Observamos en la siguiente tabla que los residuales no están incorrelados, pues muchos son lejanos a cero y por ello se salen de las bandas deseadas.

Comprobación de autocorrelación del ruido blanco									
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones					
6	436.75	6	<.0001	0.895	0.800	0.831	0.880	0.779	0.691
12	748.13	12	<.0001	0.717	0.756	0.661	0.586	0.617	0.655
18	931.54	18	<.0001	0.559	0.487	0.515	0.549	0.457	0.389
24	1044.41	24	<.0001	0.416	0.447	0.369	0.310	0.334	0.362

Tabla 18 : Residuales y comprobación de autocorrelación de la serie

Para continuar con el método ARIMA ajustado, hacemos en primer lugar una diferenciación estacional del orden 4, tomando en cuenta que nuestra serie es trimestral. Observando sus correlogramas en la figura 41, vemos que la media pasa a ser más constante y cercana a 0, los coeficientes de correlación (ACF) si bien disminuyen exponencialmente, varios pasan a ser negativos. Al fijarnos en los coeficientes de correlación parcial, la mayoría se encuentran dentro de las bandas, aunque se observa que la primera y cuarta barra aún están fuera de los rangos deseados. Todo esto nos da una pauta importante para ajustar el mejor modelo ARIMA que cumpla con las condiciones y haga que todos los residuos estén incorrelados.

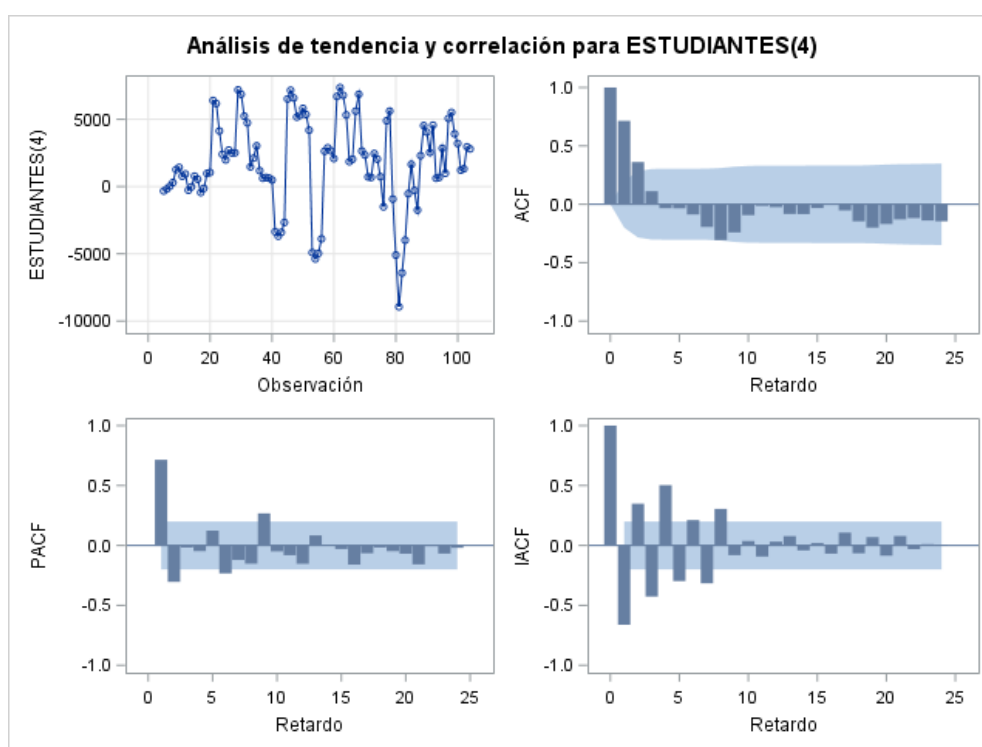


Figura41: Resultados de la serie diferenciada estacionalmente

Aplicaremos entonces un Modelo ARIMA ESTACIONAL de la forma $(p,d,q) (P,D,Q)_s$ donde aumentamos el componente d pues vamos a trabajar sobre nuestra serie diferenciada. Por tanto, el modelo ARIMA más adecuado para realizar la predicción de la serie va a ser un modelo diferenciado en la parte regular con 2 autorregresivos y una media móvil en la parte estacional. El modelo final se expresaría de la siguiente manera, **ARIMA (2,0,0) (0,1,1)₄**, mostrando los siguientes resultados.

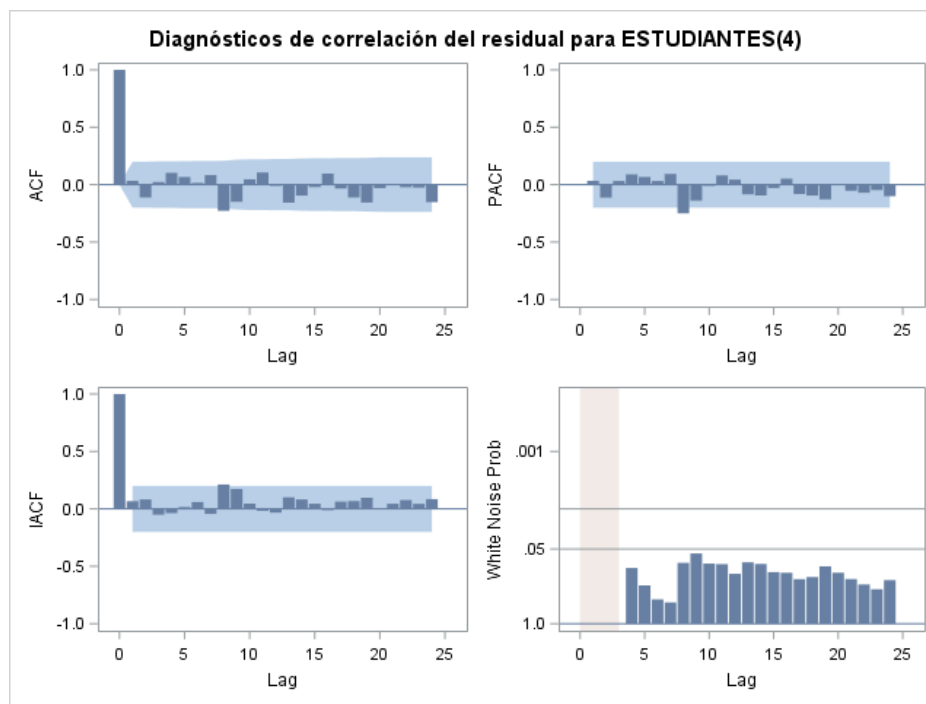


Figura 42: Representación de resultados del modelo ARIMA ajustado

Tanto en los autocorrelogramas simple como en el parcial, no se sale ninguna barra de las bandas, por lo que el ajuste en la parte regular (no estacional) $P=2$ y de $Q=(4)$ en la parte estacional ha sido el correcto, lo que nos permitirá hacer mejores predicciones sobre nuestra serie. Tras aplicar el modelo ajustado adecuado, en la siguiente tabla se puede confirmar que los residuales están incorrelados. Nuestro modelo **ARIMA (2,0,0) (0,1,1)₄** tiene los siguientes resultados. Los coeficientes de autocorrelación son todos muy cercanos a 0, lo que afirma que los residuos están incorrelados y en el gráfico simple y parcial estarán dentro de las bandas.

Comprobación de autocorrelación de residuales									
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones					
6	3.10	3	0.3767	0.033	-0.112	0.023	0.103	0.067	0.015
12	13.65	9	0.1353	0.083	-0.228	-0.149	0.047	0.106	-0.011
18	20.48	15	0.1543	-0.157	-0.095	-0.021	0.097	-0.033	-0.112
24	26.86	21	0.1754	-0.155	-0.029	0.001	-0.022	-0.025	-0.152

Estimación por mínimos cuadrados condicional						Constante Estimación	
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t	Retardo	Varianza Estimación	
MU	1751.8	412.93005	4.24	<.0001	0	Error Std Estimación	2151.909
MA1,1	0.47176	0.09570	4.93	<.0001	4	AIC	1822.528
AR1,1	0.94467	0.09965	9.48	<.0001	1	SBC	1832.948
AR1,2	-0.22136	0.10298	-2.15	0.0341	2	Número de residuales	100

Figura 43: Resultados del Modelo ARIMA (2,0,0) (0,1,1)₄

❖ Expresión algebraica del modelo

Después de observar que los p-valores para contrastar si los parámetros del modelo son distintos de cero, son todos menores a 0,05 y por tanto rechazar la hipótesis nula, aceptamos este modelo ARIMA para realizar nuestras predicciones, la expresión algebraica del mismo se obtiene en función de los factores autorregresivos y de media móvil obtenidos.

Factores autorregresivos	
Factor 1:	1 - 0.94467 B**(1) + 0.22136 B**(2)
Factores de la media móvil	
Factor 1:	1 - 0.47176 B**(4)

Figura 44: Factores autorregresivos y de media móvil del modelo

❖ La expresión algebraica de nuestro modelo **ARIMA (2,0,0) (0,1,1)₄**, sería la siguiente:

$$\begin{aligned}
 & (1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{Ps}) (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) (1 - B^s)^D (1 - B^d) X_t = \\
 & (1 - \theta_1 B - \theta_2 B^{2s} - \dots - \theta_q B^{Qs}) (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) Z_t \\
 & (1 - 0,94467B + 0,22136B^2) ((1 - B^4)X_t - \hat{\mu}) = (1 - 0,47176B^4)Z_t \\
 & (1 - 0,94467B + 0,22136B^2 - B^4 + 0,94467B^5 - 0,22136B^6)X_t - (1 - 0,94467 + 0,22136)1751.8 = (1 - 0,47176B^4)Z_t \\
 & X_t - 0,94467X_{t-1} + 0,22136X_{t-2} - X_{t-4} + 0,94467X_{t-5} - 0,22136X_{t-6} - 484.714 = Z_t - 0,47176Z_{t-4} \\
 & X_t = 484.714 + 0,94467X_{t-1} - 0,22136X_{t-2} + X_{t-4} - 0,94467X_{t-5} + 0,22136X_{t-6} - 0,47176Z_{t-4} + Z_t
 \end{aligned}$$

❖ Cálculo de predicciones con modelo ARIMA (2,0,0) (0,1,1)₄

Ahora bien, podemos obtener la predicción de estudiantes extranjeros que obtienen la autorización de estancia por estudios en España a través de nuestro modelo ARIMA ajustado, obteniendo resultados muy positivos, una parte de estos se muestra en la siguiente tabla. Fijándonos en los valores residuales, decimos que a simple vista la predicción es acertada, pues es muy cercana a los datos reales y vemos que los residuales son en general bajos.

Predicciones para la variable ESTUDIANTES					
Obs	Predicción	Error Std	Límites de confianza al 95%		Real
					Residual
5	11001.8282	2151.9093	6784.1635	15219.4929	8917.0000
6	9939.2148	2151.9093	5721.5501	14156.8794	9969.5538
7	9734.3078	2151.9093	5516.6431	13951.9725	9379.7350
8	7885.6254	2151.9093	3667.9608	12103.2901	7636.5206
9	10665.0614	2151.9093	6447.3967	14882.7261	10179.0000
10	11565.1684	2151.9093	7347.5037	15782.8331	11390.6980

Tabla 19: Predicciones del modelo ARIMA ajustado

El resultado se puede observar en el siguiente gráfico que muestra cómo se acopla nuestro modelo a la serie real y las predicciones a partir de la línea entrecortada en la parte derecha.

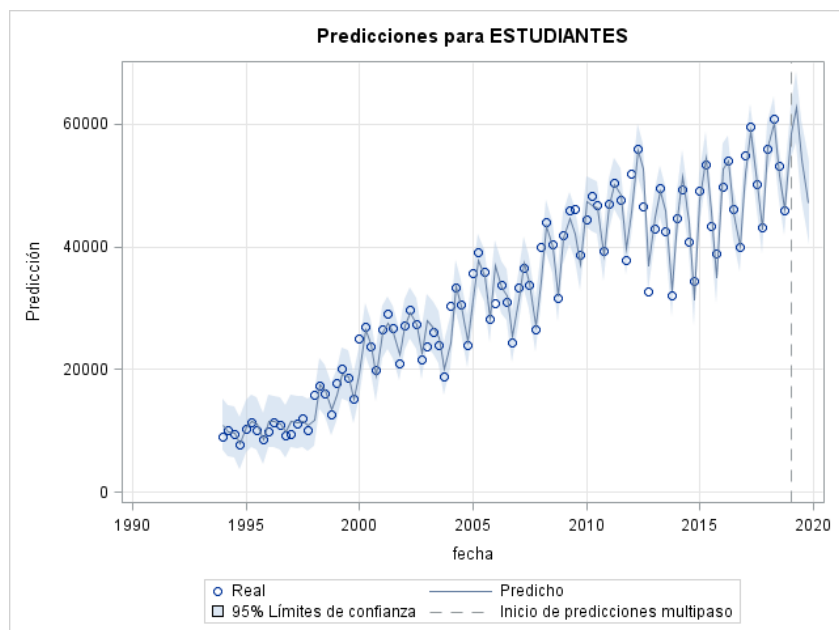


Figura 45: Representación de predicción del modelo ARIMA ajustado

En la siguiente tabla, obtenemos las predicciones para el 2019 de estudiantes extranjeros en España de nuestro modelo ARIMA ajustado.

Obs	FORECAST_ARIMA	fecha	STD	L95	U95	RESIDUAL
1	58458.93	01JAN19	2151.91	54241.26	62676.59	.
2	62688.12	01APR19	2960.26	56886.11	68490.13	.
3	54187.14	01JUL19	3293.68	47731.64	60642.64	.
4	47092.85	01OCT19	3418.18	40393.33	53792.37	.

Tabla 20: Predicciones para el 2019 del modelo ARIMA ajustado

5.4.3. Comparación y elección del mejor modelo

Después de obtener nuestras proyecciones para nuestra serie, usando el modelo de SUAVIZADO HOLT-WINTERS y luego el MODELO ARIMA AJUSTADO, procedemos a comparar estas predicciones con los datos reales de estudiantes extranjeros en España en los distintos trimestres del año 2019, datos que los habíamos separado previamente al análisis. En la siguiente figura, vemos la representación gráfica en la línea azul de los datos reales

(ESTUDIANTES) , versus las predicciones de nuestro modelo ARIMA ajustado (línea verde) y las del modelo de suavizado múltiple de Holt-Winters (línea roja).

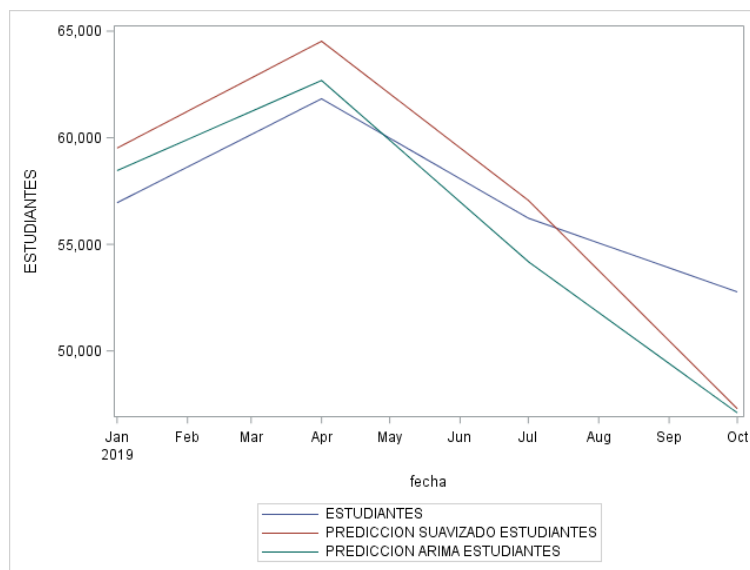


Figura 46: Representación de predicciones versus datos reales

En el primer trimestre del año las predicciones de ambos modelos son bastante cercanas a los datos reales, sin embargo, a partir del segundo trimestre, desde abril, las predicciones de ambos modelos se entrecruzan con los datos reales, por lo que la manera de comprobar cuál modelo es mejor, se hará mediante la comparación del error cuadrático medio de cada uno junto con medidas de ajuste de cada modelo como son el AIC y el SBC, observamos los resultados en la siguiente tabla.

MODELO	AIC	SBC	Variable	N	Media	Dev std
HOLT_WINTERS	1603.83	1611.76	ERROR2_SUAVIZADO	4	11139968.98	12923091.68
ARIMA_AJUSTADO	1822.53	1832.95	ERROR2_ARIMA	4	9846401.01	14972980.04

Tabla 21: Comparación del error y medidas de bondad de los modelos

A la hora de elegir el mejor modelo, si tenemos en cuenta las medidas de bondad de ajuste sobre los datos utilizados para la estimación, es mejor modelo el de suavizado porque tiene un AIC y SBC menor. Sin embargo, observando el comportamiento en los datos que habíamos reservado para probar el modelo sería mejor el ARIMA. Dado que el comportamiento de ambos modelos es muy similar comparando también la amplitud de los intervalos de predicción haremos la predicción del 2020 con los dos modelos.

5.4.4. Predicción de estudiantes extranjeros en 2020 con ARIMA

Siendo uno de los principales objetivos el de predecir el número de estudiantes extranjeros en España en 2020, aplicaremos el mejor modelo obtenido en si tomamos en cuenta el que tiene un menor error con relación a los datos reales reservados, el modelo **ARIMA (2,0,0) (0,1,1)₄**, obteniendo los siguientes resultados.

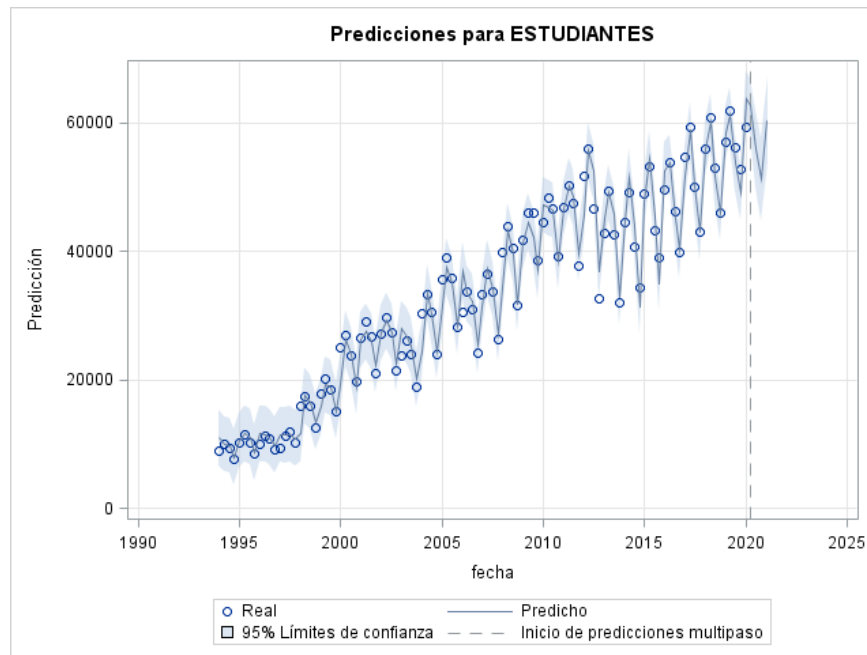


Figura 47: Representación de predicción 2020 con ARIMA ajustado

Analizando la figura 47, vemos en esta la representación de la predicción y el modelamiento entre la serie predicha representada por la línea azul y la serie real hasta 2019 representada por círculos azules, donde se corrobora que se dejan fuera muy pocas observaciones. La línea entrecortada nos muestra a su derecha la predicción para el 2020, donde se observa que la serie es bastante similar a otros años siguiendo una tendencia y estacionalidad muy parecida. A continuación, observamos los datos predichos por el modelo ARIMA. Para finales de 2020 habría un total de 60466 estudiantes extranjeros en España, es decir, **un 2% más que 2019**.

Sistema SAS					
Obs	PREDICCION_2020	fecha	STD	L95	U95
1	62732.21	01APR20	2195.49	58429.13	67035.28
2	55988.00	01JUL20	2991.57	50124.63	61851.38
3	51303.32	01OCT20	3296.64	44842.02	57764.62
4	60466.77	01JAN21	3397.83	53807.14	67126.41

Tabla 22: Predicción trimestral 2020 aplicando Arima ajustado

5.4.5. Predicción de estudiantes extranjeros en 2020 con Holt Winters

En segundo lugar, se han realizado las predicciones utilizando el modelo de suavizado de Holt-Winters, que presenta medidas de bondad como AIC y SBC menores y por lo tanto sería el mejor modelo si tomamos en cuenta estos resultados. A continuación, vemos su representación, donde se observa que el suavizado también se acopla muy bien a la serie real, dejando fuera pocas observaciones y teniendo una predicción muy cercana en tendencia y estacionalidad que la de periodos anteriores.

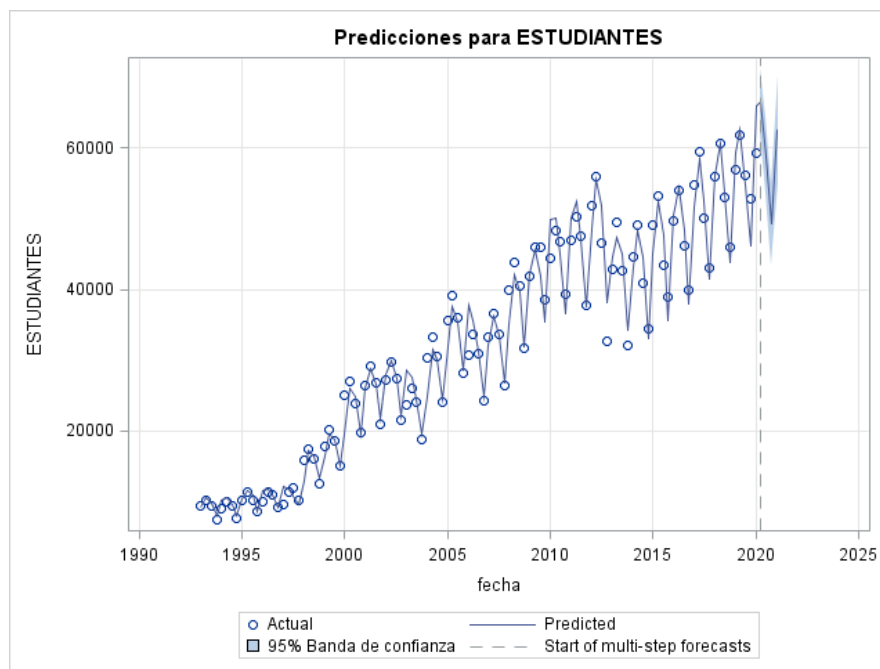


Figura 48: Representación de predicción 2020 con Holt-Winters

Las predicciones realizadas por el modelo de suavizado son las siguientes. Si comparamos estos datos con los resultados del ARIMA, el suavizado predice mayor cantidad de estudiantes extranjeros en medio, habiendo mas en el primero, segundo y cuarto semestre. Solo en el tercer semestre predice que habrá menos estudiantes que la predicción del Arima ajustado. Para concluir, decimos que con Holt-Winters, se predice que para finales de 2020 habría un total de 62620 estudiantes extranjeros en España, es decir, **un 5,64% más que 2019**.

Predicciones para la variable ESTUDIANTES					
Obs	Tiempo	Predicciones	Error estándar	Límites de confianza al 95%	
110	2020:2	66645.3465	2362.8523	62014.2410	71276.4519
111	2020:3	59014.2191	2767.0244	53590.9509	64437.4873
112	2020:4	49239.1836	2974.6007	43409.0733	55069.2939
113	2021:1	62620.7439	3882.2876	55011.6001	70229.8878

Tabla 23: Predicción trimestral 2020 aplicando Holt Winters

5.4.6. Combinación de modelos ARIMA ajustado y Holt Winters

Después de observar que ambos modelos planteados dan como resultado predicciones cercanas y sabiendo que, al comparar las predicciones de 2019 con los datos reales, nuestro modelo Arima ajustado era más cercano y tenía menor error, mientras que el de Suavizado tenía medidas de ajuste menores. Hemos decidido hacer una combinación de ambos modelos y de sus predicciones, para mejorar nuestros resultados. Esto, no es nada más que obtener la media entre estos dos y así mejorar nuestras predicciones para 2020, en la figura 49 se representa gráficamente esta combinación y cada uno de los modelos.

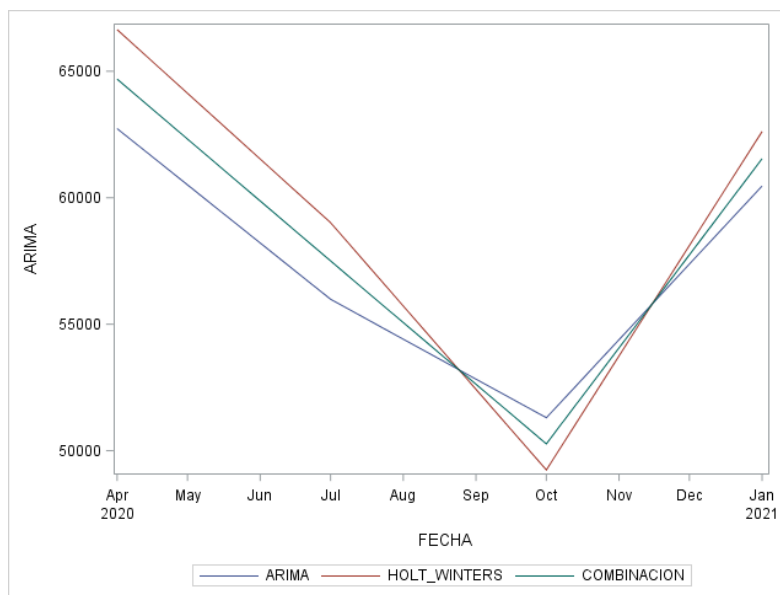


Figura 49: Representación de predicciones 2020 y combinación de modelos.

Las predicciones obtenidas con cada modelo y las obtenidas con la combinación de ambos se muestran a continuación, al tener límites de confianza bastante cercanos vemos que las predicciones de Arima y Holt_Winters son parecidas, por lo que la combinación de ambas viene a ser de utilidad para tener un resultado más equilibrado y mejorar los estadísticos de la predicción, disminuyendo el AIC Y SBC y el error de la predicción.

Obs	FECHA	ARIMA	HOLT_WINTERS	COMBINACION
1	01APR20	62732	66645	64689
2	01JUL20	55988	59014	57501
3	01OCT20	51303	49239	50271
4	01JAN21	60466	62620	61543

Tabla 24: Predicciones para 2020 y combinación de modelos

Tomando en cuenta la combinación de ambos modelos podemos concluir diciendo que para el final del año 2020 habrá 61.543 estudiantes extranjeros en España, exactamente 2.268 estudiantes más que el año anterior. Observamos también que la tendencia en las tres predicciones por trimestre se cumple, teniendo un mayor número de estudiantes extranjeros entre enero y abril, un total de 64.689 , que disminuye a 57.501 estudiantes para julio, dato que sigue disminuyendo aún más, hasta octubre donde habrá 50.271 estudiantes. En definitiva, decimos que en España **a finales del 2020 habrá un 3,8% más de estudiantes extranjeros.**

Si traducimos esta tendencia al total de estudiantes extranjeros, incluyendo no solo a los que necesitan estancia por estudios sino a todos los extracomunitarios, habrá 77.385 estudiantes extranjeros, y si tomamos en cuenta también a los comunitarios, **en España para finales de 2020 habrá aproximadamente 117.670 estudiantes extranjeros.** Datos que confirman el crecimiento continuo y que sin duda alguna representan nuevos retos y oportunidades para las instituciones que son parte del Sistema de Educación Universitario Español. A modo de resumen, se muestra en la figura 50 una infografía con las principales características de los estudiantes extranjeros en las universidades españolas en 2020.

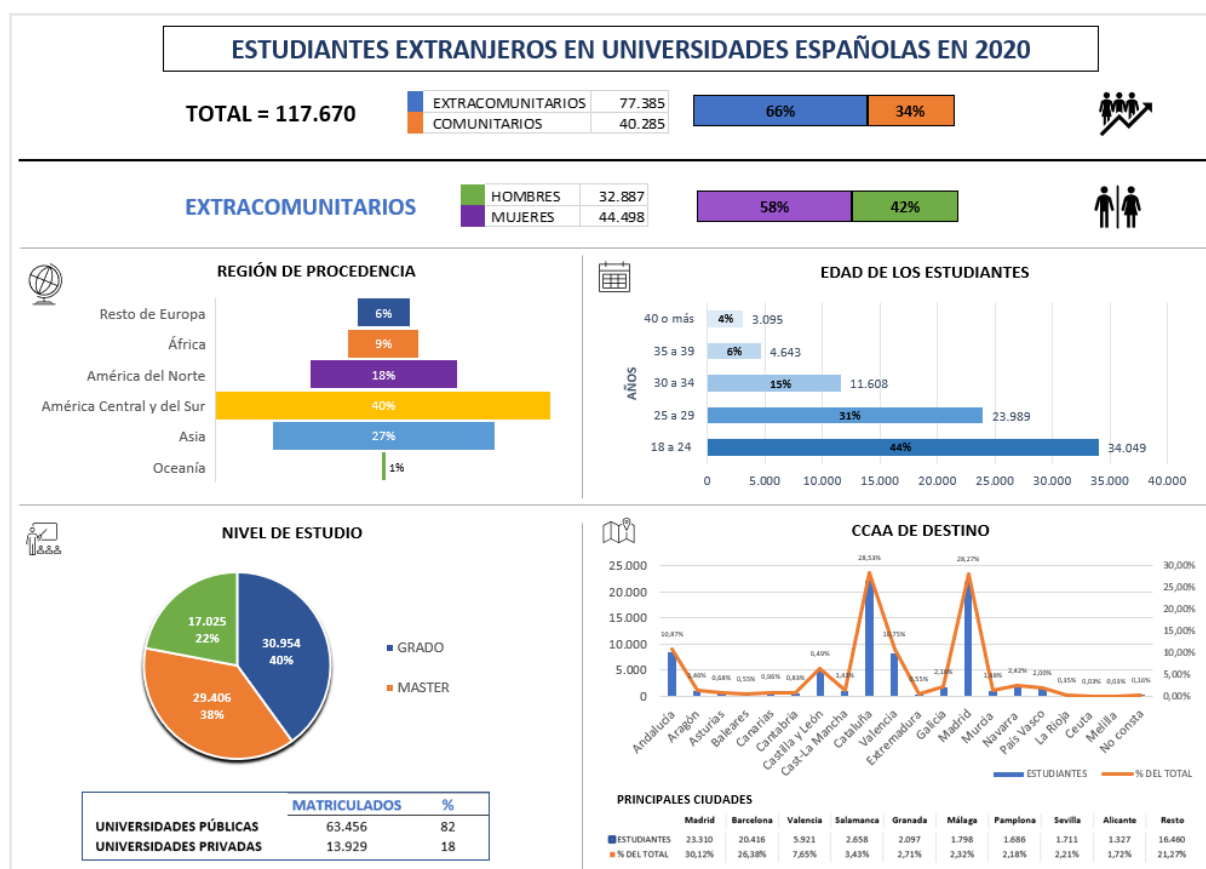


Figura50. Contexto general de estudiantes extranjeros en España en 2020

Capítulo 6

6. Conclusiones

En este apartado daremos respuesta a los objetivos planteados inicialmente incluyendo los resultados obtenidos. Recordemos que el objetivo principal de este proyecto era predecir a partir de una serie temporal, el número de estudiantes extranjeros que estudiarán en España en el año 2020 y tener un concepto claro del escenario y contexto de los estudiantes extranjeros en este país, analizando factores y variables que influyen en la elección de su destino y a la par observando la influencia que genera en España, su movilización. A partir del cumplimiento satisfactorio de este objetivo, se establecen estas conclusiones:

- ❖ España es un foco de atracción de nuevos estudiantes internacionales, efectivamente existe un continuo aumento de estudiantes extranjeros que eligen este país para realizar sus estudios universitarios.
- ❖ La movilización de estudiantes extranjeros a España ha tenido un crecimiento constante y una tendencia positiva desde hace más de 20 años, lo que demuestra la hegemonía y la ventaja competitiva que tiene el Sistema Universitario Español para captar estudiantes de distintas regiones del mundo. Esto se ha visto favorecido por la globalización y la continua intención de las Universidades de captar estudiantes extranjeros, lo que crea una comunidad multicultural y genera beneficios económicos importantes al país.
- ❖ La mayoría de los estudiantes extranjeros universitarios son mujeres, siendo en el último año (2019) un 14,41% más que los hombres. América Central y del Sur siempre ha sido la región que más estudiantes ha movilizado hacia el país, representando en promedio el 53% del total.
- ❖ En los últimos años destaca el gran aumento de estudiantes procedentes de Asia y África, principalmente de China y Marruecos, debido a acuerdos bilaterales entre estos países y España. América del norte es otra de las regiones importantes de donde han llegado una proporción (18%) de estudiantes considerable y muy estable en los últimos 17 años.
- ❖ A través de la aplicación de regresiones lineales que han hecho frente a la multicolinealidad de nuestros datos, decimos que las principales variables que influyen sobre la cantidad de estudiantes extranjeros en España son en este orden: el número de Universidad públicas y Universidades Privadas en una zona geográfica determinada, el precio de alquiler de una habitación en un piso compartido y finalmente el precio por crédito del máster habilitante

- ❖ Los análisis de correspondencias también han permitido extraer conclusiones importantes. Los principales resultados obtenidos a partir de estos son que la mayoría de los estudiantes extranjeros tienen entre 18 y 24 años, representando el 53% del total.
- ❖ Existe una gran preferencia por estudiar en Madrid y Cataluña, acogiendo cada una al 28% del total de estudiantes extranjeros en 2019. Así mismo, Andalucía y Valencia son las siguientes comunidades en este aspecto, acogiendo cada una al 11% de estudiantes. Es decir, entre estas cuatro, se concentra casi el 80% de estudiantes extranjeros en el país.
- ❖ Finalmente, tras el análisis de nuestra serie temporal, hemos obtenido un error menor aplicando un modelo Arima ajustado de la forma $ARIMA(2,0,0)(0,1,1)_4$, sin embargo, si tenemos en cuenta las medidas de bondad de ajuste sobre los datos utilizados para la estimación es mejor el suavizado de Holt Winters, teniendo un AIC y SBC menores. Por este motivo se ha decidido combinar y obtener las medias de ambos modelos, obteniendo de esta manera una predicción para el 2020 más robusta.
- ❖ Para finales de 2020 se estima que habrá un **3,8% más de estudiantes extranjeros** que solicitan autorización de estancia por estudios en España. Si traducimos esta tendencia al total de estudiantes extranjeros tomando en cuenta también a los comunitarios, **en España para finales de 2020 habrá aproximadamente 117.670 estudiantes extranjeros.**

6.1. Trabajo futuro

- ❖ Como trabajo futuro, no podemos dejar de lado los temas de actualidad en el mundo y principalmente los estragos políticos, sociales y económicos que ha causado mundialmente el Coronavirus. Como ampliación de este trabajo podría plantearse un estudio de los escenarios futuros que permita analizar las consecuencias de esta pandemia para el Sistema Universitario Español y para las personas que pretendían realizar sus estudios universitarios en España, además de los desafíos que tendrán las Universidades para cubrir esas plazas generalmente cubiertas por extranjeros y que permitían generar ingresos mayores.
- ❖ Como segundo punto, sería muy interesante aplicar métodos y algoritmos de predicción más avanzados para tener predicciones mucho más acertadas y con menor error. Para lo cual sería ideal a futuro lograr obtener el acceso a más datos y a ficheros actualizados del Sistema Universitario Español o del Ministerio de Educación, pues si bien esto ha sido una limitación para la realización de este trabajo, afortunadamente no ha llegado a ser un impedimento para cumplir con los objetivos planteados.

Bibliografía

- [1] Alonso Revenga, J.M. Análisis Descriptivo de una serie temporal. Material de la asignatura Complementos de Formación en Técnicas de Minería de Datos, Universidad Complutense de Madrid, 2016.
- [2] Alonso Revenga, J.M. 2018. Análisis de Correspondencias. Material de la asignatura Complementos de Formación en Técnicas de Minería de Datos, Universidad Complutense de Madrid, 2016.
- [3] Box,G, Jenkins,G. *Time series analysis: forecasting and control*, Holden Day, San Francisco, 1976.
- [4] Brooks, R., y Waters, J. (2011). *Student Mobilities, Migration and the Internationalization of Higher Education*. Basingstoke: Palgrave Macmillan.
- [6] Comisión Europea Eurostat. (2015). *Methodological manual on learning mobility in tertiary education*. Recuperado de <https://circabc.europa.eu/>.
- [7] De la Fuente, S. (2011).Análisis de Correspondencias Simples y Múltiples. Universidad Autónoma de Madrid. Madrid.
- [8] España, el país de la OCDE con mayor aumento de estudiantes internacionales al quitar la Selectividad a los extranjeros. (2018). *El Mundo*. Recuperado de <https://www.elmundo.es/>
- [9] Fotocasa. (6 de septiembre de 2018). Alquilar una habitación en España cuesta de media 325 euros al mes. Fotocasa. Recuperado de <http://www.fotocasa.es/>
- [10] Hernández, J., y Pérez, J. (2019). *La universidad española en Cifras: Resumen Ejecutivo de CRUE Universidades Españolas*. Recuperado de <http://www.crue.org/>
- [11] Hastie,T., Tibshirami, R., Friedman, J. (2008). *The Elements of Statistical learning*. Springer, New York.

- [12] Instituto de Estadística de la UNESCO. (2013). *Clasificación Internacional Normalizada de la Educación* (CINE). Recuperado de <https://doi.org/>
- [13] Valencia, J.L., Diaz-Llanos F.J. y Sainz-Calleja, F.J.(2003). *La regresión PLS en las ciencias experimentales*. Editorial Complutense S.A. , Madrid.
- [14] Los universitarios pagan el doble en Madrid que en Andalucía y Galicia DE. (10 de diciembre de 2019). *La Vanguardia*. Recuperado de <https://www.lavanguardia.com/>
- [15] Mauricio, J.A. *Introducción al análisis de series temporales*, Universidad Complutense de Madrid, 2007.
- [16] Ministerio de Ciencia, Innovación y Universidades. (2019). *Estadísticas de internacionalización*. Recuperado de <http://www.ciencia.gob.es/>
- [17] Ministerio de Ciencia, Innovación y Universidades. (2019). *Estadística de precios públicos universitarios*. Recuperado de <http://www.ciencia.gob.es/>
- [18] Ministerio de Ciencia, Innovación y Universidades. (2019). *Datos y cifras del Sistema Universitario Español*. Recuperado de <http://www.educacionyfp.gob.es/>
- [19] Ministerio de Inclusión, Seguridad Social y Migraciones. (2019). *Extranjeros con autorización de estancia por estudios en vigor*. Resultados. Recuperado de <http://extranjeros.mitramiss.gob.es/>
- [20] Nadal, J. O. (2016). *España como destino de estudiantes universitarios internacionales: datos y tendencias*. Universidad Politécnica de Catalunya, Barcelona. Recuperado de <https://upcommons.upc.edu/>.
- [22] Peña, D. *Análisis de series temporales*, Alianza Editorial, Madrid, 2005.

Anexos

A. Código SAS

A.1. Análisis de Correspondencias Simple

A.1.1. Análisis entre Edad y Región de procedencia

```
/* INGRESO Y MUESTRA DE DATOS */
DATA TFM2.CORRESPONDENCIA2;
INPUT EDAD $ RESTO_EUROPA AFRICA NORTEAMERICA
CENTROYSURAMERICA ASIA OCEANIA;
CARDS;
18-24 1764 3466 4656 5723 9532 131
25-29 653 1143 3115 9173 3341 79
30-34 360 461 1159 5249 1281 30
35-39 170 220 443 1822 536 11
40-mas 120 163 374 1375 372 24
RUN;

PROC PRINT DATA=TFM2.CORRESPONDENCIA2;
RUN;

/*ANÁLISIS CORRESPONDENCIAS Y DIVISIÓN FICHEROS*/
PROC CORRESP DATA=TFM2.CORRESPONDENCIA2 ALL CHI2P PRINT=BOTH;
VAR RESTO_EUROPA AFRICA NORTEAMERICA CENTROYSURAMERICA ASIA
OCEANIA;
ID EDAD;
ODS OUTPUT CELLCHISQ=TFM2.APORTACIONES;
ODS OUTPUT ROWPROFILES=TFM2.PERFILFILA;
ODS OUTPUT COLPROFILES=TFM2.PERFILCOLUMNA;
ODS OUTPUT INERTIACHART=TFM2.INERCIA;
RUN;

/*ANÁLISIS PERFIL COLUMNA*/
PROC SGPLOT DATA=TFM2.PERFILCOLUMNA;
SERIES X=LABEL Y=RESTO_EUROPA / LINEATTRS= (THICKNESS=3);
SERIES X=LABEL Y=AFRICA / LINEATTRS= (THICKNESS=3);
SERIES X=LABEL Y=NORTEAMERICA / LINEATTRS= (THICKNESS=3);
SERIES X=LABEL Y=CENTROYSURAMERICA / LINEATTRS= (THICKNESS=3);
SERIES X=LABEL Y=ASIA / LINEATTRS= (THICKNESS=3);
SERIES X=LABEL Y=OCEANIA / LINEATTRS= (THICKNESS=3);
YAXIS LABEL='Proporción';
XAXIS LABEL='Región';
TITLE "PERFIL COLUMNA";
run;
```

```

/*ANÁLISIS PERFIL FILA Y TRASPONER LABEL DE PERFIL FILA*/
PROC TRANSPOSE DATA=TFM2.PERFILFILA OUT=TFM2.PERFILFILAT;
ID LABEL;
RUN;

/* GRÁFICO PERFIL FILA */
PROC SGPLOT DATA=TFM2.PERFILFILAT;
SERIES X=_NAME_ Y=_18N24/ LINEATTRS=(THICKNESS=3);
SERIES X=_NAME_ Y=_25N29/ LINEATTRS=(THICKNESS=3);
SERIES X=_NAME_ Y=_30N34/ LINEATTRS=(THICKNESS=3);
SERIES X=_NAME_ Y=_35N39/ LINEATTRS=(THICKNESS=3);
SERIES X=_NAME_ Y=_40_mas/ LINEATTRS=(THICKNESS=3);
YAXIS LABEL='PROPORCION';
XAXIS LABEL='EDAD';
TITLE "PERFIL FILA";
RUN;

/* MAPA DE CALOR
CREACIÓN FICHERO SIN LABEL DE COLUMNA Y FILA "SUMA"*/
DATA TFM2.APORACIONES2(DROP=SUMA);
SET TFM2.APORACIONES;
IF Label="Suma" THEN DELETE;
RUN;

/* EDICIÓN FICHERO APORTACIONES PARA MAPA DE CALOR */
DATA TFM2.APORACIONES3(KEEP=FILAS COL FF);
ARRAY VECTOR {6} RESTO_EUROPA AFRICA NORTEAMERICA
CENTROYSURAMERICA ASIA OCEANIA;
SET TFM2.APORACIONES2;
A=0;
DO AUX= 'RESTO_EUROPA', 'AFRICA', 'NORTEAMERICA',
'CENTROYSURAMERICA', 'ASIA', 'OCEANIA';
A=A+1;
FILAS=LABEL;
COL=AUX;
FF=VECTOR {a};
OUTPUT;
END;
RUN;

/* REPRESENTACIÓN DEL MAPA DE CALOR */
PROC SGPLOT DATA=TFM2.APORACIONES3;
HEATMAP X=FILAS Y=COL / FREQ=FF COLORMODEL=TwoColorRamp
OUTLINE;
TITLE "APORACIONES A CHI2";
RUN;

```


A.1.2. Análisis entre Región de procedencia y CCAA de destino

```
/* INGRESO Y MUESTRA DE DATOS */
DATA TFM.CORRESPONDENCIA1;
INPUT CCAA $ RESTO_EUROPA AFRICA NORTEAMERICA
CENTROYSURAMERICA ASIA OCEANIA;
CARDS;
ANDALUCÍA 514 1021 610 1825 1579 25
ARAGÓN 32 161 64 248 319 4
ASTYCANT 25 43 64 410 68 5
CANARYBALE 78 100 55 229 130 3
CAST_LEÓN 64 171 213 1012 1361 10
CAST_LAMANCHA 22 63 50 196 229 1
CATALUÑA 1497 1017 1772 7867 4547 41
C.VALENCIANA 605 619 481 2441 1618 10
EXTREMADURA 8 41 132 125 17 6
GALICIA 44 74 105 361 375 1
MADRID 413 1129 2211 8324 5918 59
MURCIA 53 94 63 307 212 1
NAVARRA 21 84 222 965 115 10
PAISVASCO 53 58 145 533 169 5
RIOJA 9 10 28 36 22 3
CEUTAYMELILLA 7 30 25 65 20 1
RUN;

PROC PRINT DATA=TFM.CORRESPONDENCIA1;
RUN;

/*ANÁLISIS CORRESPONDENCIAS Y DIVISIÓN FICHEROS*/
PROC CORRESP DATA= TFM.CORRESPONDENCIA1 CHI2P PRINT=BOTH ALL;
VAR RESTO_EUROPA AFRICA NORTEAMERICA CENTROYSURAMERICA ASIA
OCEANIA;
ID CCAA;
ODS OUTPUT CELLCHISQ=TFM.APORTACIONES;
ODS OUTPUT ROWPROFILES=TFM.PERFILFILA;
ODS OUTPUT COLPROFILES=TFM.PERFILCOLUMNA;
ODS OUTPUT INERTIACHART=TFM.INERCIA;
RUN;

/*ANÁLISIS CORRESPONDENCIAS CON 3 DIMENSIONES*/
PROC CORRESP DATA= TFM.CORRESPONDENCIA1 DIMENS=3 ALL CHI2P
PRINT=BOTH;
VAR RESTO_EUROPA AFRICA NORTEAMERICA CENTROYSURAMERICA ASIA
OCEANIA;
ID CCAA;
ODS OUTPUT CELLCHISQ= TFM.APORTACIONES;
ODS OUTPUT ROWPROFILES=TFM.PERFILFILA;
ODS OUTPUT COLPROFILES=TFM.PERFILCOLUMNA;
ODS OUTPUT INERTIACHART=TFM.INERCIA;
RUN;
```

```

/*ANÁLISIS PERFIL COLUMNA*/
PROC SGPLOT DATA=TFM.PERFILCOLUMNA;
SERIES X=LABEL Y=RESTO_EUROPA / LINEATTRS= (THICKNESS=3);
SERIES X=LABEL Y=AFRICA / LINEATTRS= (THICKNESS=3);
SERIES X=LABEL Y=NORTEAMERICA / LINEATTRS= (THICKNESS=3);
SERIES X=LABEL Y=CENTROYSURAMERICA / LINEATTRS= (THICKNESS=3);
SERIES X=LABEL Y=ASIA / LINEATTRS= (THICKNESS=3);
SERIES X=LABEL Y=OCEANIA / LINEATTRS= (THICKNESS=3);
YAXIS LABEL='Proporción';
XAXIS LABEL='Región';
TITLE "PERFIL COLUMNA";
run;

/*ANÁLISIS PERFIL FILA
TRASPONER LABEL DE PERFIL FILA*/
PROC TRANSPOSE DATA=TFM.PERFILFILA OUT=TFM.PERFILFILAT;
ID LABEL;
RUN;

/* GRÁFICO PERFIL FILA */
PROC SGPLOT DATA=TFM.PERFILFILAT;
SERIES X=_NAME_ Y=ANDALUC_ / LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=ARAG_N/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=ASTYCANT/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=CANARYBA/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=CAST_LE_ / LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=CAST_LAM/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=CATALU_A/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=C_VALENC/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=EXTREMAD/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=GALICIA/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=MADRID/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=MURCIA/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=NAVARRA/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=PAISVASC/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=RIOJA/ LINEATTRS= (THICKNESS=3);
SERIES X=_NAME_ Y=CEUTAYME/ LINEATTRS= (THICKNESS=3);
YAXIS LABEL='PROPORCION';
XAXIS LABEL='COMUNIDAD AUTÓNOMA';
TITLE "PERFIL FILA";
RUN;

/* MAPA DE CALOR
FICHERO SIN LABEL DE COLUMNA Y FILA "SUMA" */
DATA TFM.APORCIONES2 (DROP=SUMA);
SET TFM.APORCIONES;
IF Label="Suma" THEN DELETE;
RUN;

/* EDICIÓN FICHERO APORCIONES PARA MAPA DE CALOR */
DATA TFM.APORCIONES3 (KEEP=FILAS COL FF);

```

```

ARRAY VECTOR {6} RESTO_EUROPA AFRICA NORTEAMERICA
CENTROYSURAMERICA ASIA_OCEANIA;
SET TFM.APORTACIONES2;
A=0;
DO AUX= 'RESTO_EUROPA', 'AFRICA', 'NORTEAMERICA',
'CENTROYSURAMERICA', 'ASIA', 'OCEANIA';
A=A+1;
FILAS=LABEL;
COL=AUX;
FF=VECTOR {a};
OUTPUT;
END;
RUN;

/* REPRESENTACIÓN MAPA DE CALOR */
PROC SGPLOT DATA=TFM.APORTACIONES3;
HEATMAP X=FILAS Y=COL/ FREQ=FF COLORMODEL=TwoColorRamp;
TITLE "APORTACIONES A CHI2";
RUN;

```

A.2. Regresión LASSO

```

/* INGRESO MANUAL DE DATOS*/
DATA TFM4.ESTUDEXT;
INPUT ESTUD U_PUBL U_PRIV GRADO M_HAB M_NOHAB DOCTOR ALQUILER;
DATALINES;
6441 9 1 62.23 77.83 57.00 60.30 235
867 1 1 75.13 108.92 60.76 218.95 245
406 1 0 59.73 107.84 96.21 200 235
326 1 0 92.56 45.4 49.73 210 390
567 2 2 45.65 69.12 46.58 203.8 263
491 1 1 67.28 67.43 44.38 206.6 239
3844 4 4 81.66 94.83 31.61 225.89 213
847 1 0 85.08 176.05 76.21 400.85 194
16911 7 4 106.96 148.21 109.01 401.12 472
6373 5 3 81.25 92.86 62.79 300 254
326 1 0 76.13 87.50 53.21 104 180
1281 3 0 30.28 33.93 32.61 200 197
16758 6 7 108.77 158.63 84.07 390 404
877 3 1 80.50 98.10 74.60 388 210
1434 2 1 101.56 102 73.65 323.35 247
1188 3 2 37.61 39.20 60.93 204.50 329
208 1 0 102.14 117.50 139.59 207.26 255
;
PROC PRINT DATA=TFM4.ESTUDEXT;
RUN;

/*CORRELACIÓN LINEAL ENTRE VARIABLES*/
PROC CORR DATA=TFM4.ESTUDEXT
PLOTS (MAXPOINTS=NONE) = (MATRIX (HISTOGRAM NVAR=ALL)) PEARSON;
VAR ESTUD U_PUBL U_PRIV GRADO M_HAB M_NOHAB DOCTOR ALQUILER;
RUN;

```

```

/*APLICANDO CRITERIO SBC*/
PROC GLMSELECT DATA=TFM4.ESTUDEXT PLOTS=ALL;
MODEL ESTUD=U_PUBL U_PRIV GRADO M_HAB M_NOHAB DOCTOR ALQUILER
/SELECTION=LASSO(STOP=NONE CHOOSE=SBC);
RUN;

/*MODELO EN EL PASO 4*/
PROC GLMSELECT DATA=TFM4.ESTUDEXT PLOTS=ALL;
MODEL ESTUD=U_PUBL U_PRIV GRADO M_HAB M_NOHAB DOCTOR ALQUILER
/SELECTION=LASSO(STEPS=4 CHOOSE=SBC) DETAILS=ALL STATS=ALL;
OUTPUT OUT=TFM4.PREDESTUD_SBC;
RUN;

/*VERIFICACIÓN DE UTILIDAD DE PREDICCIÓN*/
DATA TFM4.PRED_SBC;
LABEL p_ESTUD=PREDICCIÓN CON SBC;
RENAME p_ESTUD=PRED_SBC;
SET TFM4.PREDESTUD_SBC;
RUN;

PROC PRINT DATA=TFM4.PREDESTUD_SBC;
RUN;

```

A.3. Regresión PLS

```

/* INGRESO MANUAL DE DATOS*/
DATA TFM8.ESTUDIANTES;
INPUT ESTGRADO ESTMASTER ESTDOCT GRADO M_HAB M_NOHAB DOCTOR
ALQUILER RESTEUROPA AFRICA AM_NORTE AM_CENTRALYSUR ASIA
OCEANIA;
DATALINES;
3537 2329 2565 62.23 77.83 57.00 60.30 235 514 1021 610 1825 1579 25
510 140 283 75.13 108.92 60.76 218.95 245 32 161 64 248 319 4
123 179 188 59.73 107.84 96.21 200 235 9 23 27 108 48 2
286 94 153 92.56 45.4 49.73 210 390 34 19 27 74 51 1
697 98 173 45.65 69.12 46.58 203.8 263 44 81 28 155 79 2
457 564 50 67.28 67.43 44.38 206.6 239 16 20 37 302 20 3
1862 2673 1196 81.66 94.83 31.61 225.89 213 64 171 213 1012 1361 10
291 179 227 85.08 176.05 76.21 400.85 194 22 6350 196 229 1
6919 8768 4489 106.96 148.21 109.01 401.12 472 1497 1017 1772 7867 4547 39
3745 2492 1970 81.25 92.86 62.79 300 254 605 619 481 2441 1618 10
99 76 126 76.13 87.50 53.21 104 180 8 41 132 125 17 6
515 359 702 30.28 33.93 32.61 200 197 44 74 105 361 375 1
7506 8495 3163 108.77 158.63 84.07 390 404 413 1129 2211 8324 5918 59
1147 355 300 80.50 98.10 74.60 388 210 53 94 63 307 212 1
1074 984 331 101.56 102 73.65 323.35 247 21 84 222 965 115 10
762 601 571 37.61 39.20 60.93 204.50 329 53 58 145 533 169 4
103 15 22 102.14 117.50 139.59 207.26 255 9 10 28 36 22 3
;

```

```

/*ENCONTRAR EL NUMERO CORRECTO DE FACTORES*/
PROC PLS DATA=TFM8.ESTUDIANTES CV=ONE CVTEST;
MODEL ESTGRADO ESTMASTER ESTDOCT= GRADO M_HAB M_NOHAB DOCTOR
ALQUILER RESTEUROPA AFRICA AM_NORTE AM_CENTRALYSUR ASIA
OCEANIA/SOLUTION;
OUTPUT OUT=TFM8.FACTORES p=EGSTD EMSTD EDSTD;
RUN;

/*CONSTRUCCIÓN DE MATRIZ DE VARIABLES ESTANDARIZADAS*/
PROC PLS DATA=TFM8.ESTUDIANTES METHOD=PLS NFAC=2 VARSS
CENSACLE DETAILS PLOTS=(PARMPROFILES CORRLOAD(NFAC=2 UNPACK));
MODEL ESTGRADO ESTMASTER ESTDOCT= GRADO M_HAB M_NOHAB DOCTOR
ALQUILER RESTEUROPA AFRICA AM_NORTE AM_CENTRALYSUR ASIA
OCEANIA;
OUTPUT OUT=TFM8.SALIDA_PLS XSCORE=T YSCORE=U STDY=EGSTD EMSTD
EDSTD
STDY=GRSTD MH_STD MNH_STD DOCSTD ALQSTD EURSTD AFRSTD
A_NORSTD A_CYSSTD ASIASTD OCEASTD ;
RUN;

PROC PRINT DATA=TFM8.SALIDA_PLS;
FORMAT YSTD EGSTD GRSTD MH_STD MNH_STD DOCSTD ALQSTD EURSTD
AFRSTD A_NORSTD A_CYSSTD ASIASTD OCEASTD T1 T2 U1 U2 5.3;
VAR EGSTD EMSTD EDSTD GRSTD MH_STD MNH_STD DOCSTD ALQSTD
EURSTD AFRSTD A_NORSTD A_CYSSTD ASIASTD OCEASTD T1 T2 U1 U2;
RUN;

/*AGREGANDO SOLUCIÓN*/
PROC PLS DATA=TFM8.ESTUDIANTES METHOD=PLS NFAC=2 VARSS
CENSACLE DETAILS;
MODEL ESTGRADO ESTMASTER ESTDOCT=GRADO M_HAB M_NOHAB DOCTOR
ALQUILER RESTEUROPA AFRICA AM_NORTE AM_CENTRALYSUR ASIA
OCEANIA/SOLUTION;
OUTPUT OUT=TFM8.SALIDA_PLS XSCORE=T YSCORE=U STDY=EGSTD EMSTD
EDSTD
STDY=GRSTD MH_STD MNH_STD DOCSTD ALQSTD EURSTD AFRSTD
A_NORSTD A_CYSSTD ASIASTD OCEASTD;
RUN;

```

A.4. Análisis y Predicción de Serie temporal

```

/* IMPORTACIÓN FICHERO Y CREACIÓN DE FECHA */
PROC IMPORT DATAFILE= "C:\Users\Usuario\Desktop\DATOS
TFM\SERIE_ESTUDIANTES.xlsx"
OUT = TFM.SERIE dbms=xlsx;
RUN;

DATA TFM.SERIE;
SET TFM.SERIE;

```

```

fecha = intnx("QTR", "01JAN1993"d, _N_-1);
FORMAT FECHA DATE.;
RUN;

proc print DATA= TFM.SERIE;
RUN;

/* ESTADÍSTICOS DESCRIPTIVOS */
PROC MEANS DATA=TFM.SERIE;
VAR ESTUDIANTES;
RUN;

/* REPRESENTACIÓN SERIE */
PROC SGPLOT DATA=TFM.SERIE;
SERIES X=FECHA Y=ESTUDIANTES;
RUN;

/* REPRESENTACIÓN SERIE 2013-2016 */
PROC SGPLOT DATA= TFM.SERIE;
WHERE FECHA BETWEEN "01JAN2013"d AND "01JAN2016"d;
SERIES X=FECHA Y=ESTUDIANTES;
RUN;

/* DESCOMPOSICIÓN ESTACIONAL */
PROC TIMESERIES DATA=TFM.SERIE PLOT=(DECOMP PERIODOGRAM
SERIES)
PRINT=(SEASONS DECOMP);
ID FECHA INTERVAL=QTR;
VAR ESTUDIANTES;
RUN;

/* DIVISIÓN DE DATOS EN DOS FICHEROS */
DATA TFM.SERIE_SE TFM.SERIE_SP;
SET TFM.SERIE;
IF FECHA<'01JAN2019'd THEN OUTPUT TFM.SERIE_SE;
ELSE OUTPUT TFM.SERIE_SP;
RUN;

PROC PRINT DATA= TFM.SERIE_SP;
RUN;

/* SUAVIZADO DE HOLT-WINTERS */
PROC ESM DATA = TFM.SERIE_SE LEAD=4 OUT=TFM.SERIE_PREDSUAV
PRINT=(FORECASTS ESTIMATES)
PLOT=(FORECASTS MODELS SEASONS);
ID FECHA INTERVAL = QTR;
FORECAST ESTUDIANTES / MODEL=MULTwinters;
RUN;

```

```

/*DIVISIÓN DE PREDICCIONES HOLT-WINTERS EN DOS FICHEROS */

DATA TFM.SERIE_PREDSUAV2019;
SET TFM.SERIE_PREDSUAV;
WHERE FECHA>='01JAN2019'd ;
RUN;

/* RENOMBRE DE LABEL PARA FUTURA COMPARACIÓN */
DATA TFM.SERIE_PREDSUAV2019;
LABEL ESTUDIANTES=PREDICCIÓN SUAVIZADO ESTUDIANTES;
RENAME ESTUDIANTES=FORECAST_SUAVIZADO;
SET TFM.SERIE_PREDSUAV2019;
RUN;

/* APLICACIÓN MODELOS ARIMA
REPRESENTACIÓN CORRELOGRAMAS */
PROC ARIMA DATA= TFM.SERIE_SE;
IDENTIFY VAR=ESTUDIANTES;
RUN;

/* DIFERENCIACIÓN ESTACIONAL DE LA SERIE */
PROC ARIMA DATA= TFM.SERIE_SE;
IDENTIFY VAR= ESTUDIANTES (4);
RUN;

/* MODELO ARIMA AJUSTADO */
PROC ARIMA DATA= TFM.SERIE_SE;
IDENTIFY VAR= ESTUDIANTES (4);
ESTIMATE P=2 Q=(4);
RUN;

/* PREDICCIÓN CON MODELO ARIMA */
PROC ARIMA DATA= TFM.SERIE_SE PLOTS=ALL;
IDENTIFY VAR= ESTUDIANTES (4);
ESTIMATE P=2 Q=(4) OUTMODEL=MODEL OUTEST=AUTO;
FORECAST LEAD=4 ID=FECHA INTERVAL=QTR OUT=
TFM.SERIE_SE_PREDARIMA PRINTALL;
RUN;

/* CREACIÓN FICHERO PREDICCIÓN CON MODELO ARIMA */
data TFM.SERIE_PREDARIMA2019;
set TFM.SERIE_SE_PREDARIMA;
WHERE ESTUDIANTES=. ;
run;

/* RENOMBRE DE LABEL PARA COMPARACIÓN */
DATA TFM.SERIE_PREDARIMA2019 (DROP=ESTUDIANTES);
LABEL FORECAST=PREDICCIÓN ARIMA ESTUDIANTES;
RENAME FORECAST=FORECAST_ARIMA;
SET TFM.SERIE_PREDARIMA2019;
RUN;

```

```

PROC PRINT DATA= TFM.SERIE_PREDARIMA2019;
RUN;

/* CREACIÓN DE FICHERO PARA COMPARACIÓN DE MODELOS */
DATA TFM.UNIONMODELOS (DROP= STD L95 U95 RESIDUAL);
MERGE TFM.SERIE_SP TFM.SERIE_PREDSUAV2019
TFM.SERIE_PREDARIMA2019;
ERROR_SUAVIZADO=ESTUDIANTES-FORECAST_SUAVIZADO;
ERROR2_SUAVIZADO=ERROR_SUAVIZADO**2;
ERROR_ARIMA=ESTUDIANTES-FORECAST_ARIMA;
ERROR2_ARIMA=ERROR_ARIMA**2;
RUN;

proc print DATA=TFM.UNIONMODELOS;
RUN;

/* REPRESENTACIÓN COMPARACIÓN DE MODELOS */
PROC SGPLOT DATA=TFM.UNIONMODELOS;
SERIES X=FECHA Y=ESTUDIANTES;
SERIES X=FECHA Y=FORECAST_SUAVIZADO;
SERIES X=FECHA Y=FORECAST_ARIMA;
RUN;

/* CUADRO DE DATOS - COMPARACIÓN DE MODELOS */
PROC PRINT DATA=TFM.UNIONMODELOS;
VAR ESTUDIANTES FORECAST_SUAVIZADO ERROR_SUAVIZADO
FORECAST_ARIMA ERROR_ARIMA;
RUN;

/* ERROR CUADRÁTICO MEDIO DE LOS MODELOS */
PROC MEANS DATA=TFM.UNIONMODELOS;
VAR ERROR2_SUAVIZADO ERROR2_ARIMA;
RUN;

/*PREDICCIÓN PARA 2020 CON MODELO ARIMA JUSTADO*/
PROC ARIMA DATA= TFM.SERIE PLOTS=ALL;
IDENTIFY VAR= ESTUDIANTES(4);
ESTIMATE P=2 Q=(4) OUTMODEL=MODEL OUTEST=AUTO;
FORECAST LEAD=4 ID=FECHA INTERVAL=QTR OUT=
TFM.SERIE_PREDARIMA2020 PRINTALL;
RUN;
data TFM.SERIE_PREDICIONES2020 (DROP=ESTUDIANTES RESIDUAL);
LABEL FORECAST=PREDICCIÓN_2020 ;
RENAME FORECAST=PREDICCIÓN_2020;
SET TFM.SERIE_PREDICIONES2020;
RUN;

PROC PRINT DATA=TFM.SERIE_PREDICIONES2020;
RUN;

```


- **Predicciones 2020 y combinación de modelos**

```

PROC IMPORT DATAFILE= "C:\Users\Usuario\Desktop\DATOS TFM\DATOS
SERIE TEMPORAL\SERIE_ESTUDIANTES.xlsx"
OUT = TFM10.SERIE dbms=xlsx;
RUN;

DATA TFM10.SERIE;
SET TFM10.SERIE;
fecha = intnx("QTR", "01JAN1993"d, _N_-1);
FORMAT FECHA DATE.;
RUN;

/*#####PREDICCIÓN PARA 2020 CON SUAVIZADO HOLT WINTERS#####*/
PROC ESM DATA = TFM10.SERIE LEAD=4 OUT=TFM10.SERIE_PREDSUAV2020
PRINT=(FORECASTS ESTIMATES)
PLOT=(FORECASTS MODELS SEASONS);
ID FECHA INTERVAL = QTR;
FORECAST ESTUDIANTES / MODEL=MULTwinters;
RUN;

/*##DIVISIÓN DE PREDICCIONES DE HOLT-WINTERS EN DOS FICHEROS###*/
DATA TFM10.SERIE_PREDSUAVHW2020;
SET TFM10.SERIE_PREDSUAV2020;
WHERE FECHA>='01JAN2020'd ;
RUN;

/*#####RENOMBRE DE LABEL PARA FUTURA COMPARACIÓN#####*/

DATA TFM10.SERIE_PREDSUAVHW2020;
LABEL ESTUDIANTES=PREDICCIÓN HOLT-WINTERS ESTUDIANTES;
RENAME ESTUDIANTES=FORECAST_HOLTWINTERS;
SET TFM10.SERIE_PREDSUAVHW2020;
RUN;

/*PREDICCIÓN PARA 2020 CON MODELO ARIMA AJUSTADO*/
PROC ARIMA DATA= TFM10.SERIE PLOTS=ALL;
IDENTIFY VAR= ESTUDIANTES(4);
ESTIMATE P=2 Q=(4) OUTMODEL=MODEL OUTEST=AUTO;
FORECAST LEAD=4 ID=FECHA INTERVAL=QTR OUT= TFM10.SERIE_PREDARIMA2020
PRINTALL;
RUN;

/*CREACIÓN FICHERO PREDICCIÓN CON MODELO ARIMA para 2020*/
data TFM10.SERIE_ARIMA2020;
set TFM10.SERIE_PREDARIMA2020;
WHERE ESTUDIANTES=. ;
run;
data TFM10.SERIE_ARIMA2020 (DROP=ESTUDIANTES RESIDUAL);
LABEL FORECAST=PREDARIMA_2020 ;
RENAME FORECAST=PREDARIMA_2020;
SET TFM10.SERIE_ARIMA2020;
RUN;

```

```

proc print DATA=TFM10.SERIE_ARIMA2020;
RUN;

/*#####COMBINACIÓN DE MODELOS#####*/
PROC IMPORT DATAFILE= "C:\Users\Usuario\Desktop\DATOS TFM\DATOS
SERIE TEMPORAL\COMBINACION.xlsx"
OUT = TFM10.COMBINACION dbms=xlsx;
RUN;

DATA TFM10.COMBINACION ;
SET TFM10.COMBINACION ;
fecha = intnx("QTR", "01APR2020"d, _N_-1);
FORMAT FECHA DATE.;
RUN;

proc print DATA=TFM10.COMBINACION;
RUN;

PROC SGPLOT DATA=TFM10.COMBINACION;
SERIES X=FECHA Y=ARIMA;
SERIES X=FECHA Y=HOLT_WINTERS;
SERIES X=FECHA Y=COMBINACION;
RUN;

```

B. Tablas de datos

En este anexo se muestran tablas de datos que fueron de suma importancia para la realización de este trabajo y para la consecución de los objetivos planteados.

- Datos de análisis de variable Sexo

AÑO	HOMBRES	MUJERES	DIFERENCIA	%
2002	11.363	12.365	1.002	4,22%
2003	14.434	15.812	1.378	4,56%
2004	16.666	18.879	2.213	6,23%
2005	14.473	16.167	1.694	5,53%
2006	15.540	17.727	2.187	6,57%
2007	18.662	21.312	2.650	6,63%
2008	19.225	22.580	3.355	8,03%
2009	20.236	24.210	3.974	8,94%
2010	20.925	25.989	5.064	10,79%
2011	22.461	29.343	6.882	13,28%
2012	19.063	23.801	4.738	11,05%
2013	19.540	24.979	5.439	12,22%
2014	21.422	27.631	6.209	12,66%
2015	21.498	28.171	6.673	13,43%
2016	23.263	31.476	8.213	15,00%
2017	23.558	32.395	8.837	15,79%
2018	24.246	32.705	8.459	14,85%
2019	25.368	33.907	8.539	14,41%
PROMEDIO	19.552	24.414	4.861	10,23%

- Datos de análisis de variable Comunidad Autónoma de destino

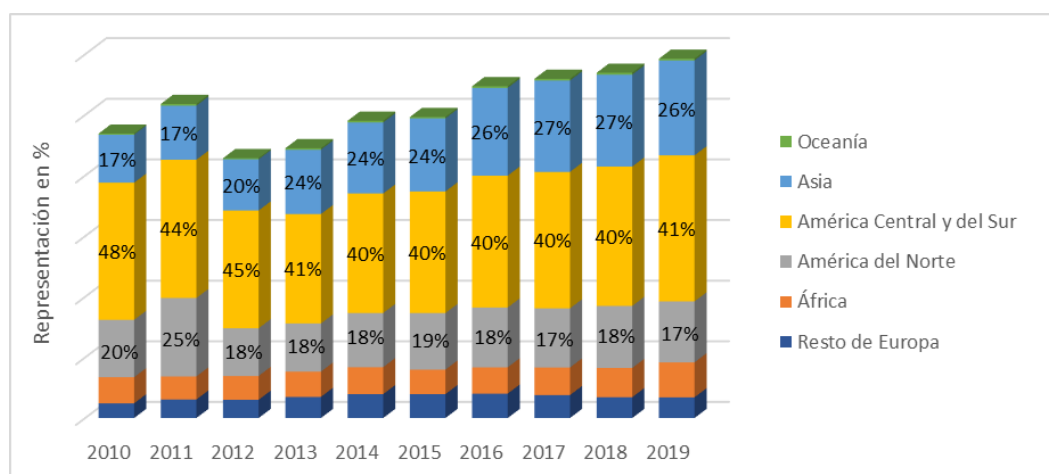
CCAA	ESTUDIANTES	% DEL TOTAL
Andalucía	6.441	10,87%
Aragón	867	1,46%
Asturias	406	0,68%
Baleares	326	0,55%
Canarias	567	0,96%
Cantabria	491	0,83%
Castilla y León	3.844	6,49%
Castilla-La Mancha	847	1,43%
Cataluña	16.911	28,53%
Valenciana	6.373	10,75%
Extremadura	326	0,55%
Galicia	1.281	2,16%
Madrid	16.758	28,27%
Murcia	877	1,48%
Navarra	1.434	2,42%
País Vasco	1.188	2,00%
La Rioja	208	0,35%
Ceuta	15	0,03%
Melilla	19	0,03%

- Datos de análisis de variable Región de procedencia

REGIÓN / AÑO	2010	2011	2012	2013	2014
Resto de Europa	2.427	3.055	3.012	3.485	3.937
África	4.303	3.807	3.923	4.163	4.445
América del Norte	9.459	12.931	7.875	7.948	8.943
América Central y del Sur	22.623	22.798	19.415	18.039	19.704
Asia	7.927	8.928	8.441	10.618	11.723
Oceanía	170	285	193	248	278
TOTAL	46.909	51.804	42.859	44.501	49.030
REGIÓN / AÑO	2015	2016	2017	2018	2019
Resto de Europa	3.939	4.023	3.776	3.432	3.412
África	4.041	4.340	4.554	4.825	5.763
América del Norte	9.332	9.869	9.756	10.229	10.057
América Central y del Sur	20.062	21.743	22.501	22.982	24.085
Asia	12.056	14.480	15.069	15.169	15.664
Oceanía	222	274	290	309	288
TOTAL	49.652	54.729	55.946	56.946	59.269

- Información ampliada de variable Región y país de procedencia

Gráfico de evolución de la proporción de estudiantes por Región de procedencia.



El auge de extranjeros procedentes de China y Marruecos

Es importante destacar que el auge de estudiantes provenientes de China se debe a dos motivos principales, el primero es el crecimiento de la clase media alta en este país que paulatinamente se interesa en realizar sus estudios en el exterior, lo cual se ve beneficiado por el segundo motivo que son los estrechos acuerdos bilaterales entre España y esta potencia asiática, principalmente el Convenio sobre Reconocimiento a efectos académicos con China¹⁰.

¹⁰ Ministerio de Ciencia, Innovación y Universidades. Aplicación provisional del Acuerdo en materia de reconocimientos de títulos y diplomas firmado en Pekín en 2007.

Un motivo muy similar al segundo mencionado anteriormente es el que hace que muchos estudiantes marroquíes se encuentren dentro del sistema universitario español, además del acuerdo de cooperación en materia cultural y educativa entre Marruecos y España, se suma el hecho del continuo aumento de migraciones principalmente de jóvenes menores de edad que son acogidos por el gobierno como refugiados y luego pasan a formar parte del sistema universitario español.

- Datos de análisis de variable Nivel de estudios

CCAA	GRADO	MASTER	DOCTORADO
Andalucía	3.537	2.329	2.565
Aragón	510	140	283
Asturias	123	179	188
Baleares	286	94	153
Canarias	697	98	173
Cantabria	457	564	50
Castilla y León	1.862	2.673	1.196
Castilla-La Mancha	291	179	227
Cataluña	6.919	8.768	4.489
Valenciana	3.745	2.492	1.970
Extremadura	99	76	126
Galicia	515	359	702
Madrid	7.506	8.495	3.163
Murcia	1.147	355	300
Navarra	1.074	984	331
País Vasco	762	601	571
La Rioja	103	15	22
TOTAL	29.633	28.401	16.509

- Datos de análisis de variable Precios

1. Precios públicos universitarios para Comunitarios y Extracomunitarios

PRECIOS PÚBLICOS UNIVERSITARIOS							
	ESPAÑOLES Y COMUNITARIOS			EXTRANJEROS EXTRACOMUNITARIOS			
	PRIMERA MATRÍCULA			TERCERA Y CUARTA MATRÍCULA			
CCAA	GRADO	MASTER HABILITANTE	MASTER NO HABILITANTE	GRADO	MASTER HABILITANTE	MASTER NO HABILITANTE	PRECIO TUTELA DOCTORADO
Andalucía	12,62 €	13,68 €	13,68 €	62,23 €	77,83 €	57,00 €	60,30 €
Aragón	18,74 €	23,02 €	37,40 €	75,13 €	108,92 €	60,76 €	218,95 €
Asturias	12,21 €	18,90 €	23,84 €	59,73 €	107,84 €	96,21 €	200,00 €
Balears	16,13 €	24,98 €	27,55 €	92,56 €	45,40 €	49,73 €	210,00 €
Canarias	11,71 €	11,80 €	13,62 €	45,65 €	69,12 €	46,58 €	203,80 €
Cantabria	13,03 €	18,73 €	27,31 €	67,28 €	67,43 €	44,38 €	206,60 €
Castilla y León	15,81 €	15,81 €	15,81 €	81,66 €	94,83 €	31,61 €	225,89 €
Castilla-La Mancha	17,38 €	31,14 €	39,50 €	85,08 €	176,05 €	76,21 €	400,85 €
Cataluña	33,52 €	41,17 €	48,18 €	106,96 €	148,21 €	109,01 €	401,12 €
C. Valencia	17,33 €	17,33 €	39,27 €	81,25 €	92,86 €	62,79 €	300,00 €
Extremadura	14,74 €	21,67 €	32,75 €	76,13 €	87,50 €	53,21 €	104,00 €
Galicia	11,89 €	11,89 €	26,49 €	30,28 €	33,93 €	32,61 €	200,00 €
Madrid	24,03 €	30,84 €	45,02 €	108,77 €	158,63 €	84,07 €	390,00 €
Murcia	15,58 €	21,80 €	39,77 €	80,50 €	98,10 €	74,60 €	388,00 €
Navarra	19,65 €	22,65 €	28,35 €	101,56 €	102,00 €	73,65 €	323,35 €
País Vasco	16,88 €	23,78 €	31,74 €	37,61 €	39,20 €	60,93 €	204,50 €
La Rioja	19,77 €	25,07 €	35,50 €	102,14 €	117,50 €	139,59 €	207,26 €
PRECIO PROMEDIO	17,12 €	22,01 €	30,93 €	76,15 €	95,61 €	67,82 €	249,68 €

2. Precio promedio de alquiler de una habitación en piso compartido por CCAA

CCAA	PRECIO ALQUILER/m2	PRECIO POR HABITACIÓN EN PISO COMPARTIDO
Andalucía	8,50 €	235,00 €
Aragón	7,80 €	245,00 €
Asturias	7,20 €	235,00 €
Baleares	12,60 €	390,00 €
Canarias	10,00 €	263,00 €
Cantabria	7,60 €	239,00 €
Castilla y León	6,50 €	213,00 €
Castilla-La Mancha	5,40 €	194,00 €
Cataluña	14,20 €	472,00 €
Comunitat Valenciana	7,50 €	254,00 €
Extremadura	5,00 €	180,00 €
Galicia	6,60 €	197,00 €
Madrid	14,70 €	404,00 €
Murcia	6,10 €	210,00 €
Navarra	8,50 €	247,00 €
País Vasco	11,90 €	329,00 €
La Rioja	6,50 €	255,00 €

- Contribuciones del Análisis de Correspondencias entre región y CCAA de destino

Contribuciones al estadístico chi-cuadrado total							
Porcentajes	RESTO_EUROPA	AFRICA	NORTEAMERICA	CENTROYSURAMERICA	ASIA	OCEANIA	Suma
ANDALUCÍ	2.019	15.192	0.003	3.932	0.082	0.056	21.282
ARAGÓN	0.160	2.798	0.196	0.898	0.503	0.014	4.570
ASTYCANT	0.099	0.033	0.006	1.598	1.667	0.101	3.505
CANARYBA	1.097	1.166	0.043	0.107	0.286	0.013	2.712
CAST_LEÓ	1.600	0.430	0.755	1.097	7.458	0.001	11.342
CAST_LAM	0.103	0.125	0.056	0.260	0.541	0.009	1.095
CATALUÑA	5.018	2.470	0.092	0.605	0.841	0.083	9.109
C.VALENC	4.134	0.870	0.923	0.131	0.127	0.099	6.284
EXTREMAD	0.170	0.151	5.786	0.070	1.545	0.518	8.239
GALICIA	0.086	0.012	0.001	0.229	0.657	0.034	1.019
MADRID	10.064	2.268	0.498	0.287	1.337	0.000	14.452
MURCIA	0.035	0.407	0.093	0.020	0.002	0.019	0.577
NAVARRA	1.156	0.236	0.618	4.171	5.150	0.142	11.472
PAISVASC	0.014	0.149	0.315	0.607	1.109	0.025	2.218
RIOJA	0.020	0.002	0.496	0.069	0.073	0.456	1.116
CEUTAYME	0.011	0.578	0.104	0.000	0.302	0.013	1.007
Suma	25.787	26.886	9.983	14.081	21.681	1.582	100.000

- Datos de la Serie temporal Principal

FECHA	ESTUDIANTES	FECHA	ESTUDIANTES	FECHA	ESTUDIANTES
01/01/1993	9.250	01/01/2002	27.099	01/01/2011	46.914
01/04/1993	10.157	01/04/2002	29.756	01/04/2011	50.314
01/07/1993	9.353	01/07/2002	27.400	01/07/2011	47.481
01/10/1993	7.334	01/10/2002	21.486	01/10/2011	37.712
01/01/1994	8.917	01/01/2003	23.737	01/01/2012	51.804
01/04/1994	9.970	01/04/2003	26.064	01/04/2012	55.918
01/07/1994	9.380	01/07/2003	24.001	01/07/2012	46.550
01/10/1994	7.637	01/10/2003	18.820	01/10/2012	32.611
01/01/1995	10.179	01/01/2004	30.253	01/01/2013	42.864
01/04/1995	11.391	01/04/2004	33.219	01/04/2013	49.492
01/07/1995	10.147	01/07/2004	30.589	01/07/2013	42.556
01/10/1995	8.566	01/10/2004	23.987	01/10/2013	32.082
01/01/1996	9.906	01/01/2005	35.545	01/01/2014	44.519
01/04/1996	11.372	01/04/2005	39.030	01/04/2014	49.202
01/07/1996	10.927	01/07/2005	35.940	01/07/2014	40.806
01/10/1996	9.115	01/10/2005	28.182	01/10/2014	34.378
01/01/1997	9.459	01/01/2006	30.640	01/01/2015	49.053
01/04/1997	11.238	01/04/2006	33.644	01/04/2015	53.254
01/07/1997	11.903	01/07/2006	30.981	01/07/2015	43.347
01/10/1997	10.167	01/10/2006	24.293	01/10/2015	38.943
01/01/1998	15.853	01/01/2007	33.267	01/01/2016	49.669
01/04/1998	17.407	01/04/2007	36.528	01/04/2016	53.923
01/07/1998	16.029	01/07/2007	33.637	01/07/2016	46.193
01/10/1998	12.569	01/10/2007	26.376	01/10/2016	39.929
01/01/1999	17.829	01/01/2008	39.974	01/01/2017	54.739
01/04/1999	20.112	01/04/2008	43.893	01/04/2017	59.427
01/07/1999	18.520	01/07/2008	40.418	01/07/2017	50.109
01/10/1999	15.078	01/10/2008	31.694	01/10/2017	43.129
01/01/2000	25.012	01/01/2009	41.829	01/01/2018	55.953
01/04/2000	26.964	01/04/2009	45.930	01/04/2018	60.745
01/07/2000	23.751	01/07/2009	46.032	01/07/2018	53.070
01/10/2000	19.812	01/10/2009	38.550	01/10/2018	45.932
01/01/2001	26.479	01/01/2010	44.465	01/01/2019	56.951
01/04/2001	29.075	01/04/2010	48.273	01/04/2019	61.828
01/07/2001	26.773	01/07/2010	46.750	01/07/2019	56.228
01/10/2001	20.994	01/10/2010	39.225	01/10/2019	52.768
				01/01/2020	59.275

- Resultados de Descomposición estacional de la Serie temporal

Descomposición estacional para la variable ESTUDIANTES						
Obs	Tiempo	Estación	Ciclo de tendencia	Estacional	Irregular	Ajustado estacionalmente
1	1993:1	1	.	1.044847	.	8852.97
2	1993:2	2	.	1.133451	.	8961.00
3	1993:3	3	8981.80	1.014685	1.026238	9217.47
4	1993:4	4	8916.76	0.807016	1.019184	9087.83
5	1994:1	1	8896.71	1.044847	0.959260	8534.26
6	1994:2	2	8937.89	1.133451	0.984097	8795.75
7	1994:3	3	9133.45	1.014685	1.012102	9243.99
8	1994:4	4	9468.85	0.807016	0.999347	9462.66
9	1995:1	1	9742.43	1.044847	0.999965	9742.09
10	1995:2	2	9954.53	1.133451	1.009547	10049.57
11	1995:3	3	10036.57	1.014685	0.996397	10000.41
12	1995:4	4	10000.17	0.807016	1.061401	10614.19

- Muestra de datos de predicciones de Modelo ARIMA ajustado

Predicciones para la variable ESTUDIANTES						
Obs	Predicción	Error Std	Límites de confianza al 95%		Real	Residual
5	11001.8282	2151.9093	6784.1635	15219.4929	8917.0000	-2084.8282
6	9939.2148	2151.9093	5721.5501	14156.8794	9969.5538	30.3390
7	9734.3078	2151.9093	5516.6431	13951.9725	9379.7350	-354.5728
8	7885.6254	2151.9093	3667.9608	12103.2901	7636.5206	-249.1048
9	10665.0614	2151.9093	6447.3967	14882.7261	10179.0000	-486.0614
10	11565.1684	2151.9093	7347.5037	15782.8331	11390.6980	-174.4704
11	11094.8792	2151.9093	6877.2145	15312.5439	10147.2666	-947.6126
12	8649.2326	2151.9093	4431.5679	12866.8972	8565.8253	-83.4073
13	11601.0055	2151.9093	7383.3408	15818.6702	9906.0000	-1695.0055
14	11494.1152	2151.9093	7276.4506	15711.7799	11372.4743	-121.6409
15	11122.2465	2151.9093	6904.5818	15339.9112	10927.1055	-195.1410
16	9830.6127	2151.9093	5612.9480	14048.2773	9114.8560	-715.7567
17	11536.3838	2151.9093	7318.7191	15754.0485	9459.0000	-2077.3838
18	11370.7736	2151.9093	7153.1089	15588.4383	11237.6609	-133.1127
19	11475.4746	2151.9093	7257.8099	15693.1392	11902.5034	427.0288
20	10888.5099	2151.9093	6670.8452	15106.1746	10166.5352	-721.9747
21	11701.3238	2151.9093	7483.6591	15918.9885	15853.0000	4151.6762
22	17592.5916	2151.9093	13374.9269	21810.2562	17407.2123	-185.3793
23	16598.5709	2151.9093	12380.9062	20816.2356	16029.2225	-569.3484
24	13524.5430	2151.9093	9306.8783	17742.2077	12569.3309	-955.2121
.....						
94	53901.4310	2151.9093	49683.7664	58119.0957	53922.6532	21.2221
95	45385.3478	2151.9093	41167.6831	49603.0125	46193.0000	807.6522
96	40071.8535	2151.9093	35854.1888	44289.5181	39928.9091	-142.9444
97	51844.0458	2151.9093	47626.3811	56061.7105	54739.0000	2894.9542
98	58968.6295	2151.9093	54750.9648	63186.2942	59426.8480	458.2185
99	50374.0423	2151.9093	46156.3777	54591.7070	50109.0000	-265.0423
100	42961.9751	2151.9093	38744.3104	47179.6398	43128.8581	166.8830
101	56014.0278	2151.9093	51796.3631	60231.6925	55953.0000	-61.0278
102	60133.8785	2151.9093	55916.2139	64351.5432	60744.8149	610.9364
103	51695.0639	2151.9093	47477.3992	55912.7286	53070.0000	1374.9361
104	46040.2648	2151.9093	41822.6001	50257.9295	45931.8177	-108.4471
105	58458.9281	2151.9093	54241.2634	62676.5928	-	-
106	62688.1219	2960.2647	56886.1098	68490.1340	-	-
107	54187.1396	3293.6842	47731.6373	60642.6419	-	-
108	47092.8503	3418.1848	40393.3312	53792.3694	-	-

*Tabla continúa en SAS